Spiking-Aided Neural Architecture for Efficient and Robust WiFi Sensing

Yisha Lu[1,3], Liwen Jing[1*], Jiangmao Zheng[2], Bowen Zhang[3*]

[1]Pengcheng Laboratory
[2]Shenzhen X-institute
[3]Shenzhen Technology University
luyisha@sztu.edu.cn, jinglw@pcl.ac.cn, zhengjiangmao@x-institute.edu.cn, zhang_bo_wen@foxmail.com

Abstract

This paper introduces a spiking-aided wifi sensing network (**SWS-Net**), a novel hybrid neural architecture that integrates Spiking Neural Networks (SNNs) with conventional Artificial Neural Networks (ANNs) for robust WiFi-based indoor sensing. WiFi signals offer a low-cost and device-free solution for recognizing human activities, gestures, identities and etc. However, their susceptibility to multipath fading and environmental noise poses significant challenges. Inspired by the human brain's capability to process noisy information, SWS-Net leverages the noise-resilient dynamics of spiking neurons alongside the feature extraction ability of ANNs. We present a theoretical analysis comparing the noise-handling capacities of SNNs and ANNs, and show how their combination yields both improved robustness and training efficiency. Experimental results across three WiFi sensing tasks demonstrate that SWS-Net consistently achieves higher accuracy and faster convergence compared to baseline models, validating its effectiveness in challenging indoor environments.

Code — https://github.com/Coralinehh/Wi-Fi.git

Introduction

WiFi sensing has emerged as a promising technology for various human-centric applications, including medical monitoring, smart homes, smart industry, human-computer interaction, security and etc. (Wei et al. 2025). Leveraging existing wireless infrastructure, it provides a cost-effective and non-intrusive solution for capturing human activities and environmental changes. One key advantage of WiFi sensing lies in its inherent respect for privacy (Li et al. 2022). Unlike image or video-based sensing, electromagnetic signals used in WiFi systems do not capture visual identities. Instead, they detect motion and signal fluctuations caused by physical movement. This makes WiFi sensing particularly appealing for privacy-sensitive environments such as hospitals, homes, and workplaces (He et al. 2024).

Recent advances in deep learning have significantly expanded the capabilities of WiFi sensing across a wide range of applications. These include human activity recognition (HAR) (Li et al. 2025a), gait-based human identification (Human-ID), gesture recognition (GR) for human-machine interaction (Wang et al. 2025b), vital sign monitoring such as respiration and heartbeat detection (Liang et al. 2024) and etc. These developments underscore the growing importance of integrating sensing functionality into existing WiFi communication infrastructure. While neural network models like Convolutional Neural Network (CNN), Long Short Term Memory networks (LSTM), transformer and etc. achieve high recognition accuracy, deploying deep learning for WiFi sensing in practical settings faces several challenges. WiFi devices, primarily designed for communication, operate with limited computational resources, necessitating neural network models with reduced computational complexity. Additionally, WiFi signals are susceptible to environmental noise and multipath propagation, which can degrade learning efficiency and slow convergence in conventional ANNs (Yang et al. 2023).

In this work, we propose a hybrid architecture that integrates spiking neural networks with artifitial neural networks to enable robust and efficient WiFi sensing. The SNN component excels at processing temporal signal sequences and suppressing noise, while the ANN component handles feature extraction and classification. This synergistic design leverages the strengths of both paradigms to achieve high accuracy, lower computational cost, and enhanced robustness in noisy environments.

The main contributions of this work are as follows:

(1) A lightweight ANN-SNN hybrid architecture SWS-Net is proposed, with modality-specific preprocessing for Channel State Information (CSI) and Body-coordinate Velocity Profile (BVP) signals. The model achieves higher recognition accuracy, reduced training time, and more stable validation performance compared to conventional approaches.

(2) A theoretical analysis of WiFi signal noise is conducted across SNN, ANN, and cascaded ANN-SNN architectures to provide insights into the noise-handling mechanisms and justify the design of the SWS-Net.

(3) Extensive experiments are performed on a diverse set of WiFi sensing tasks, including human activity recognition, human identification, and gesture recognition. The results demonstrate superior performance of the proposed method compared to state-of-the-art approaches.

*Corresponding author.

Related Work

ANN-Based WiFi Sensing

In indoor environments, radio frequency (RF) technologies have demonstrated significant potential for sensing applications. Among these, WiFi sensing stands out due to its widespread deployment and accessibility. Early research has explored the use of WiFi signal measurements—such as Received Signal Strength Indicator (RSSI) and CSI—for a broad range of sensing tasks. These include human presence detection (Zou et al. 2017a), human activity recognition (Zou et al. 2019; Yang et al. 2022, 2018), human identification (Zou et al. 2018; Yang et al. 2022), people counting (Zou et al. 2017b; Fan et al. 2024), and vital sign monitoring.

To tackle these complex tasks, deep learning models have been increasingly adopted, replacing traditional model-based approaches and delivering superior performance. Commonly used models include Multi-Layer Perceptrons (MLPs) (Liu, Zhao, and Chen 2017), CNNs (Moshiri et al. 2021), Recurrent Neural Networks (RNNs) (Moshiri et al. 2021), LSTM networks (Zhang et al. 2025), Gated Recurrent Unit (GRUs) (Zheng et al. 2019), and Transformer architectures (Li et al. 2021). Generative AI has also been applied to enable WiFi-based imaging (Shi et al. 2024). These models are particularly effective at processing time-series data and extracting discriminative features from complex signal patterns.

However, WiFi signal propagation in indoor environments is often affected by severe multipath interference, causing significant variability across different spatial configurations. To address this domain variability, Zheng et al. (Zheng et al. 2019) proposed transforming CSI into the BVP domain for zero-effort domain adaptation. Similarly, Zhao et al. (Zhao et al. 2025) introduced a Siamese network-based approach for cross-domain sensing with limited input samples. Moreover, He et al. (He et al. 2024) proposed a joint communication and sensing framework to reduce interference and overhead, laying the foundation for future standardization efforts such as IEEE 802.11bf.

While existing research has largely focused on improving feature extraction for specific sensing tasks, handling noise in WiFi signals and improving training efficiency remain critical challenges. In this work, we propose a hybrid model that integrates SNNs with traditional ANNs, aiming to enhance noise robustness, accelerate convergence, and improve generalization in WiFi sensing tasks.

Robustness of Noise Handling in SNN

SNNs have gained attention for their energy-efficient and biologically inspired architectures, especially under noisy or resource-constrained environments. Recent studies have demonstrated that SNNs exhibit inherent robustness to noise due to their temporal encoding and event-driven processing mechanisms. For instance, SNNs built on small-world or scale-free topologies have shown improved resistance to impulse noise, attributed to high clustering and efficient synaptic propagation structures (Guo et al. 2022, 2023). These network configurations enhance the fault tolerance of SNNs by promoting redundant and short communication paths among neurons.

In the context of signal denoising, SNNs have been successfully applied to real-world sensing modalities. A recent work proposed a noise-injected spiking graph convolution framework for 3D point cloud denoising, combining the sparsity of spike-based representation with graph-based locality (Li et al. 2025b). Similarly, spike-based image denoising techniques have shown that Leaky Integrate-and-Fire (LIF) neurons can effectively suppress Gaussian noise, with coding schemes such as rate and temporal coding contributing to noise-resilient signal reconstruction (Castagnetti, Pegatoquet, and Miramond 2023). Moreover, dynamic synapses and stochastic neural modulation have been explored as optimization tools to improve weak signal detection under noise, leveraging stochastic resonance and temporal integration (Castagnetti, Pegatoquet, and Miramond 2023).

These findings highlight the natural compatibility of SNNs with noisy input scenarios, making them a promising candidate for robust sensing tasks such as WiFi CSI analysis, where multipath fading and signal distortion are prevalent. However, integrating SNNs into practical sensing pipelines often requires hybrid designs and task-specific optimizations, which motivates this work's exploration of ANN–SNN hybrid architectures for robust and efficient WiFi sensing.

WiFi Noise Handling Analysis in ANN and SNN Architectures

A comparative analysis is conducted on the noise resilience of artificial neural networks and spiking neural networks in the context of processing Wi-Fi CSI. The noisy Wi-Fi CSI signal is modeled as:

$$H(f,t) = \underbrace{\sum_{l=1}^{L} \alpha_l(f,t) e^{-j2\pi f \tau_l(f,t)} \cdot e^{j\epsilon(f,t)}}_{\text{clean signal } s(f,t)} + n_a(t), \quad (1)$$

where $s(f,t)$ is the deterministic multi-path signal component. The model incorporates two primary noise sources: a multiplicative phase jitter, modeled by $e^{j\epsilon(f,t)}$ where the phase error $\epsilon(f,t)$ is a random variable drawn from $U(-\delta,\delta)$, and an additive white Gaussian noise (AWGN) term, $n_a(t) \sim \mathcal{N}(0,\sigma^2)$. We assume the phase jitter is small ($\delta \ll 1$), allowing for linearization. Using a first-order Taylor approximation, $e^{j\epsilon} \approx 1 + j\epsilon$, the signal model can be rewritten as:

$$H(f,t) \approx s(f,t) + \underbrace{j\epsilon(f,t)s(f,t)}_{\text{phase noise } n_p(f,t)} + \underbrace{n_a(t)}_{\text{additive noise}} . \quad (2)$$

This approximation, valid for small δ, transforms the model into a sum of the clean signal, a signal-dependent phase noise term n_p, and a signal-independent additive noise term n_a. The error of this approximation is on the order of $\mathcal{O}(\delta^2)$.

ANN Response to Composite Noise

Assuming a typical ANN with convolutional neural network (CNN) processing, the input $H(f,t)$ is processed by a convolutional layer followed by a ReLU activation function:

$$\mathbf{z}_{\text{out}} = \text{ReLU}(\mathbf{W} * H + \mathbf{b}), \tag{3}$$

where $\mathbf{W}$ is the convolution kernel and $\mathbf{b}$ is the bias. Due to the linearity of convolution, the pre-activation value can be decomposed:

$$\mathbf{z}_{\text{pre}} = \underbrace{(\mathbf{W} * s + \mathbf{b})}_{\text{effective signal } \mathbf{s}'} + \underbrace{(\mathbf{W} * n_p + \mathbf{W} * n_a)}_{\text{effective noise } \mathbf{n}'}. \tag{4}$$

The effective noise $\mathbf{n}'$ is a composite of filtered phase noise and filtered AWGN. For analytical tractability, we approximate $\mathbf{n}'$ as a zero-mean Gaussian process, $\mathbf{n}' \sim \mathcal{N}(0, \sigma^2_{\mathbf{W}})$. This is justifiable under the central limit theorem if the kernel $\mathbf{W}$ has a large receptive field. The total noise variance is $\sigma^2_{\mathbf{W}} = \text{Var}(\mathbf{W} * n_p) + \text{Var}(\mathbf{W} * n_a)$. Approximating the filtered signal power, this becomes $\sigma^2_{\mathbf{W}} \approx \frac{\delta^2}{3}\|\mathbf{W} * s\|^2_F + \sigma^2\|\mathbf{W}\|^2_F$, where $\|\cdot\|_F$ is the Frobenius norm.

SNN Response via Integrate-and-Fire Neuron

SNNs process information using discrete-time neuron models. A common choice is the Integrate-and-Fire (IF) neuron, whose behavior at each timestep t is described by a three-step process: charge (integration), fire, and reset, as shown in the provided diagram. First, the charge or integration step updates the membrane potential. The neuron integrates an input current $I(t)$ with its previous potential $V(t-1)$ to produce a new pre-activation potential $H(t)$:

$$H(t) = V(t-1) + I(t) \tag{5}$$

Here, we use a simple non-leaky integrator model. The input current $I(t)$ is derived from the magnitude of the CSI signal, $I(t) = |H(f,t)|$, as the neuron processes signal intensity. Second, in the fire step, the neuron decides whether to emit a spike $S(t)$ by comparing its pre-activation potential $H(t)$ to a fixed threshold V_{th}:

$$S(t) = \Theta(H(t) - V_{\text{th}}), \quad \text{where} \quad \Theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{6}$$

where $\Theta(\cdot)$ is the Heaviside step function. Third, in the reset step, the membrane potential $V(t)$ for the next timestep is determined. Aligning with the "Soft Reset" mechanism, the potential is updated as follows:

$$V(t) = H(t) - S(t) \cdot V_{\text{th}} \tag{7}$$

If the neuron fires ($S(t) = 1$), its potential is reduced by the threshold value. If it does not fire ($S(t) = 0$), its potential remains at the integrated value, $V(t) = H(t)$, carrying it over to the next integration step. The core strengths of this model are its thresholding mechanism and temporal integration. For the purpose of signal-to-noise ratio (SNR) analysis over a decision window of T timesteps, we analyze the accumulated potential just before the firing decision at time T. In this regime, we assume the neuron starts from a resting state $V(0) = 0$ and has not fired yet. The pre-activation potential $H(T)$ is then the direct sum of all input currents up to that point:

$$H(T) = \sum_{t=1}^{T} I(t) \sim \mathcal{N}\left(T\mu_s, T\sigma^2_{|n'|}\right). \tag{8}$$

Here, $\mu_s = |s|$ is the mean magnitude of the clean signal component, and $\sigma^2_{|n'|}$ is the variance of the magnitude of the total effective noise. This temporal summation enhances the signal mean linearly with time ($T\mu_s$) while the noise standard deviation grows more slowly, with $\sqrt{T}$. The statistics of this accumulated potential $H(T)$ are what determine the probability of a correct spike, forming the basis of our SNR analysis.

Comparative SNR Analysis of ANN vs. SNN

To quantitatively compare noise resilience, we analyze the SNR at the output of each network architecture. The key distinction lies in how each architecture processes the input and affects the final SNR.

Output SNR of an ANN The ANN processes the input through a linear convolution followed by a non-linear ReLU activation. The input SNR to the ReLU activation for a single element is:

$$\text{SNR}_{\text{ANN,in}} = \frac{(s')^2}{\sigma^2_W} \tag{9}$$

where s' is the effective signal and σ^2_W is the variance of the effective noise after convolution. The output of the ReLU for an input element $z = s' + n'$ is a random variable. We can define the output SNR as the ratio of the squared mean of the output to its variance:

$$\text{SNR}_{\text{ANN,out}} = \frac{(\text{E}[\text{ReLU}(z)])^2}{\text{Var}(\text{ReLU}(z))} \tag{10}$$

where the mean and variance are given by the moments of a rectified Gaussian distribution:

$$\text{E}[\text{ReLU}(z)] = s'\Phi\left(\frac{s'}{\sigma_W}\right) + \sigma_W\phi\left(\frac{s'}{\sigma_W}\right) \tag{11}$$

$$\begin{aligned}\text{Var}(\text{ReLU}(z)) =&((s')^2 + \sigma^2_W)\Phi\left(\frac{s'}{\sigma_W}\right) + s'\sigma_W\phi\left(\frac{s'}{\sigma_W}\right) \\ &- (\text{E}[\text{ReLU}(z)])^2\end{aligned} \tag{12}$$

Here, ϕ and Φ are the standard normal PDF and CDF, respectively. The behavior of this output SNR is characterized by high input SNR and low input SNR cases. When $s' \gg \sigma_W$), $\text{ReLU}(z) \approx z$. The activation is nearly linear, and the output SNR is largely preserved: $\text{SNR}_{\text{ANN,out}} \approx \text{SNR}_{\text{ANN,in}}$. When $s' \to 0$, noise is rectified, producing a non-zero DC offset: $\text{E}[\text{ReLU}(n')] = \sigma_W/\sqrt{2\pi}$. Any weak signal must compete against this noise-induced floor. The output signal power $(\text{E}[\text{ReLU}(z)])^2$ becomes very small, while the output variance remains significant, causing a drastic degradation of the output SNR.

Furthermore, the multi-layer noise propagation model, $\text{Var}(z^{(L)}) \approx \sigma^2 \prod_{\ell=1}^{L} \|\mathbf{W}^{(\ell)}\|^2_F$, suggests noise variance

can be amplified. In summary, the ANN tends to propagate and amplify noise, with rectification significantly degrading the SNR, especially in low-signal conditions.

Output SNR of an SNN The SNN's resilience stems from temporal integration and thresholding. The input SNR to the neuron is $\text{SNR}_{\text{SNN,in}} = \frac{\mu_s^2}{\sigma_{|n'|}^2}$. The SNR of the integrated membrane potential over T timesteps (before firing) is:

$$\text{SNR}_{\text{SNN,integrated}} = \frac{(T\mu_s)^2}{T\sigma_{|n'|}^2} = T \cdot \text{SNR}_{\text{SNN,in}} \tag{13}$$

A more meaningful output SNR for the event-based SNN is the ratio of the probability of a correct spike (signal present) to a false spike (noise only). Firing Probability with Signal (P_{S+N}):

$$P_{S+N} = P(H(T) \geq V_{\text{th}}) = 1 - \Phi\left(\frac{V_{\text{th}} - T\mu_s}{\sqrt{T}\sigma_{|n'|}}\right) \tag{14}$$

Firing Probability with Noise Only (P_N):

$$P_N = P\left(\sum_{t=1}^{T} |n'(t)| \geq V_{\text{th}}\right) = 1 - \Phi\left(\frac{V_{\text{th}}}{\sqrt{T}\sigma_{|n'|}}\right) \tag{15}$$

The output SNR of the SNN is therefore:

$$\text{SNR}_{\text{SNN,out}} = \frac{P_{S+N}}{P_N} = \frac{1 - \Phi\left(\frac{V_{\text{th}} - T\mu_s}{\sqrt{T}\sigma_{|n'|}}\right)}{1 - \Phi\left(\frac{V_{\text{th}}}{\sqrt{T}\sigma_{|n'|}}\right)} \tag{16}$$

This demonstrates that by setting a suitable threshold V_{th}, the false alarm probability P_N can be made vanishingly small, while temporal integration ensures P_{S+N} remains high. The SNN acts as a highly effective noise gate, suppressing sub-threshold noise while enhancing the signal SNR via integration, leading to a high output SNR.

Enhanced Noise Filtering in an ANN-SNN Cascade

In a hybrid architecture, the ANN's output, $z_{\text{ANN}}(t)$, becomes the input current for the SNN stage: $I(t) = |z_{\text{ANN}}(t)| = s_{\text{eff}}(t) + n_{\text{eff}}(t)$. Here, $s_{\text{eff}}(t) \approx \mu_{\text{eff}}$ is the enhanced signal magnitude from the ANN, and $n_{\text{eff}}(t) \sim \mathcal{N}(0, \sigma_{\text{eff}}^2)$ is the residual noise. The input SNR to the SNN stage is:

$$\text{SNR}_{\text{in}} = \frac{\mu_{\text{eff}}^2}{\sigma_{\text{eff}}^2} \tag{17}$$

The SNN integrates this current over T timesteps. The accumulated potential $V(T)$ follows a new Gaussian distribution:

$$V(T) \sim \mathcal{N}(T\mu_{\text{eff}}, T\sigma_{\text{eff}}^2) \tag{18}$$

The output SNR of the entire cascaded system, $\text{SNR}_{\text{cascade}}$, is the ratio of firing probabilities:

$$\text{SNR}_{\text{cascade}} = \frac{P_{S+N}}{P_N} = \frac{1 - \Phi\left(\frac{V_{\text{th}} - T\mu_{\text{eff}}}{\sqrt{T}\sigma_{\text{eff}}}\right)}{1 - \Phi\left(\frac{V_{\text{th}}}{\sqrt{T}\sigma_{\text{eff}}}\right)} \tag{19}$$

The synergy is clear: the ANN performs feature-domain filtering, and the subsequent SNN stage performs spatio-temporal filtering, using temporal integration to further boost the SNR and a threshold to gate the residual noise. This makes the hybrid architecture exceptionally robust.

Model Architecture Design

System Overview

A hybrid architecture that integrates ANNs with SNNs is proposed, aiming to achieve efficient and robust WiFi sensing in different tasks. As shown in Fig. 1, this architecture is designed with two separate network tailored to two WiFi data modalities of both CSI and BVP data.

For CSI format data, we first extract spatial features using two Spiking-Conv Blocks (SCBs) driven by Integrate-and-Fire (IF) neurons. The first SCB, equipped with 5×5 convolutional kernels, captures coarse-grained spatial patterns, while the second SCB uses 3×3 kernels to further extract fine-grained features. IF neurons are incorporated into each layer to simulate the firing mechanism of biological neurons, thereby enhancing the model's ability to perceive dynamic spatiotemporal variations. The extracted features are then flattened and passed into the Spiking Classifier Block, which consists of two linear layers. The first layer preserves the spiking dynamics to enhance nonlinear representation capacity, while the second layer performs conventional linear classification to produce the final multi-class outputs.

For BVP format data, we apply a single SCB with a 3×3 convolutional kernel to extract spatial features, considering that BVP signals inherently carry compact and abstract motion representations. To maintain a lightweight design, we omit additional convolutional layers. The extracted features are then flattened and passed into a Spiking-GRU Block, which includes two linear layers and a GRU module. The first linear layer compresses the high-dimensional feature maps into a lower-dimensional representation suitable for GRU input, while preserving the spiking characteristics through an IF neuron. The GRU module then effectively captures short-term temporal dynamics in the BVP sequence. Finally, the second linear layer serves as the classifier, recognizing temporal patterns and producing the final multi-class predictions.

Input Processing The CSI and BVP data have different spatiotemporal properties and require separate preprocessing strategies. For CSI, dynamic activities such as walking or falling induce short-term signal fluctuations. To capture these local motion patterns and spatial variations, we divide each CSI sample into $T = 10$ uniform time steps along the temporal axis. In contrast, BVP is derived from CSI using compressed sensing and beamforming velocity projection, inherently encoding global motion dynamics. As gestures often span the entire sequence, we retain the complete BVP signal. To ensure uniform input size, we adopt the maximum sequence length $t_{\text{max}} = 38$ observed across all samples, and pad shorter sequences with zeros.

Spiking-Conv Block A unified SCB is employed in this study as the core spatial feature extraction module. This

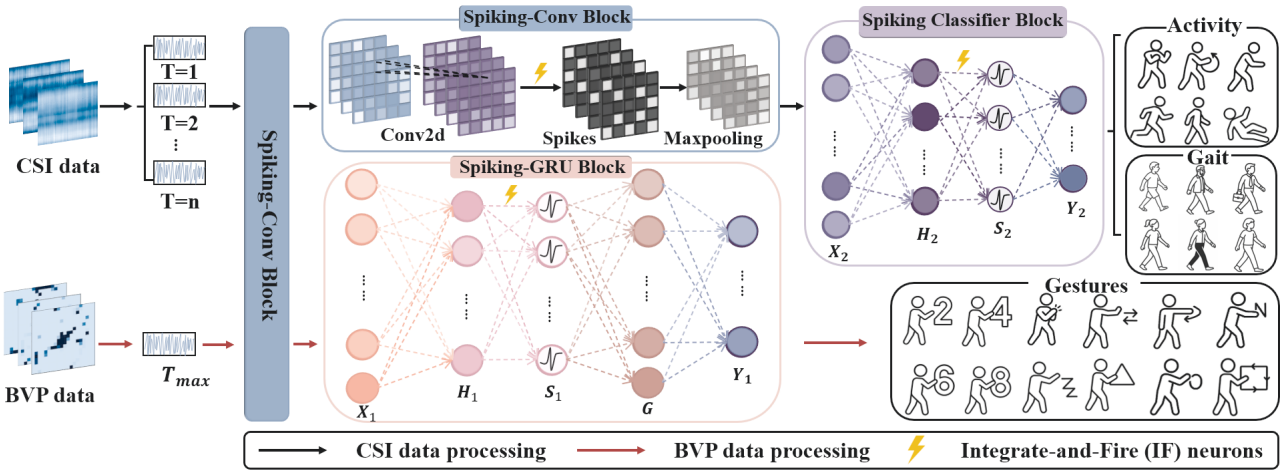


Figure 1: Framework of the hybrid SNN and ANN model SWS-Net for WiFi sensing

module is adopted in the front-end processing stages of both CSI and BVP modalities. Each SCB consists of four standard operations: a 2D convolutional layer (Conv2D), batch normalization (BatchNorm), an IF neuron with an ATan surrogate function, and a max pooling layer. In each SCB, local spatial structure features are first extracted through convolution. The output is then normalized via batch normalization to accelerate convergence and improve training stability. Subsequently, the spiking neuron encodes information in the temporal domain. Finally, max pooling is applied for spatial downsampling, reducing the dimensionality of the feature maps while preserving the most salient response regions.

In the CSI processing module, we employ two consecutive SCBs to extract both coarse-grained and fine-grained spatial features from the high-dimensional input data. This design enables the network to capture the complex spatial dependencies resulting from multipath effects and body-induced perturbations in CSI signals.

In contrast, only one SCB is used in the BVP processing module. Since the BVP data has already undergone compression and projection operations that preserve essential motion dynamics, a single SCB is sufficient for extracting meaningful spatial representations. This choice also helps reduce the network's complexity and ensures lightweight processing for gesture recognition tasks.

Spiking-GRU Block To transform the spatial feature maps into temporally discriminative representations, we introduce a Spiking-GRU Block that integrates feature compression, spiking nonlinearity, recurrent modeling, and final classification.

Specifically, the spatial features X_1 produced by the preceding Spiking Convolutional Block are first flattened into a 1D vector, followed by a linear projection layer to reduce dimensionality and generate compact temporal embeddings H_1 for each frame. Next, a spiking IF neuron with an ATan surrogate function introduces nonlinearity and temporal modulation, resulting in spike-encoded representations S_1.These embeddings are then passed into a GRU network denoted as G, which captures sequential dependencies across time. Finally, the output hidden state from the GRU is fed into a linear layer to produce the final gesture prediction logits Y_1.

Spiking Classifier Block To perform classification based on the extracted spatial features from the CSI data, we design a Spiking Classifier Block, composed of a flattening operation and two linear layers, with a spiking neuron module inserted in between.

Specifically, the spatial feature maps X_2 generated by the final SCB module are first flattened into a one-dimensional vector. This vector is passed through a latent projection layer to obtain compressed hidden representations H_2. A spiking IF neuron with an ATan surrogate function is then applied to introduce nonlinearity and temporal modulation, resulting in spike-based representations S_2. Finally, a second linear layer produces the output logits Y_2 corresponding to the target gesture classes.

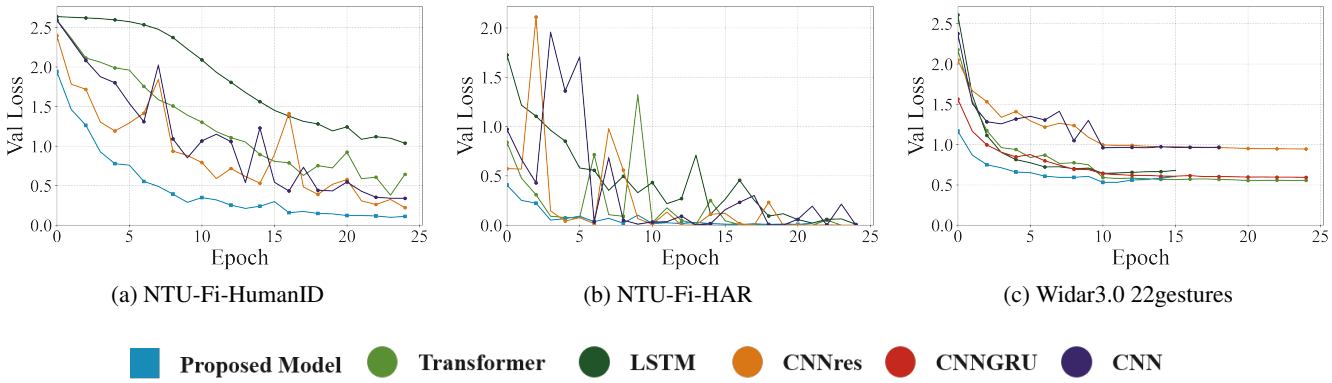
Experiments

Experimental Setup

Dataset Description We evaluate our model on three publicly available WiFi-based sensing datasets: NTU-Fi-HAR (Yang et al. 2022), NTU-Fi-HumanID (Wang et al. 2022), and Widar3.0 (Zhang et al. 2021). These datasets span three distinct recognition tasks, including human activity recognition, user identification, and gesture recognition, allowing us to comprehensively assess the generalization and robustness of the proposed hybrid spiking model.

NTU-Fi-HAR is a dataset for human activity recognition. Each sample contains CSI data of shape $3 \times 114 \times 500$, representing 3 antenna pairs, 114 subcarriers, and 500 time steps.It includes 6 activity classes: box, circle, clean, fall, run, and walk, with 936 training and 264 test samples.

NTU-Fi-HumanID focuses on identifying individuals based on walking patterns. Each sample has the same format as NTU-Fi-HAR. The dataset includes gait data from 14 subjects, with 546 training and 294 test samples.



(a) NTU-Fi-HumanID (b) NTU-Fi-HAR (c) Widar3.0 22gestures

Figure 2: Validation loss of the SWS-Net versus baselines on NTU-Fi-HumanID, NTU-Fi-HAR and Widar3.0 datasets.

The Widar3.0 dataset comprises 43,527 WiFi sensing samples covering 22 human gestures, including basic actions (push/pull, sweep, clap, slide), symbolic motions (Draw-N/O/Rectangle/Triangle/Zigzag), and numeric gestures (Draw-0 to Draw-9). The data is split into training (81% with 10% validation) and test (10%) sets. Notably, the Widar dataset encompasses recordings collected across diverse environments and under varying signal conditions, implicitly covering a range of SNR levels.

Baselines and Criterion To evaluate the effectiveness of our proposed hybrid spiking model, we conduct extensive experiments across three representative WiFi sensing tasks: human activity recognition (NTU-Fi-HAR), user identification (NTU-Fi-HumanID), and gesture recognition (Widar3.0 22-gestures). We compare our model against several baselines, including CNN, CNNRes, LSTM, Transformer, and CNNGRU. The CNN and CNNRes architectures extract spatial features using convolutional layers, with the latter incorporating residual connections for improved feature learning. LSTM and Transformer models integrate convolutional frontends with temporal modules to capture sequential dependencies, using recurrent units or self-attention respectively. CNNGRU, applied specifically to BVP data, combines frame-wise CNN feature extraction with GRU-based temporal modeling. It is included as an ablation baseline to isolate the contribution of the spiking component in our SWS-Net. To further validate this, we replaced the spiking (IF) neurons with ReLU activations, and the resulting model achieved about 80% accuracy—nearly identical to the CNN-GRU baseline (also ReLU), about 3–4% lower than our SNN-GRU. Therefore, we only report CNN-GRU as the Widar ablation in the main results.

To enable a fair comparison with baseline models, we adopt the following evaluation metrics:

Accuracy (%):Average classification accuracy across test samples, defined as:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(\hat{y}_i = y_i) \times 100 \tag{20}$$

where N is the total number of test samples, y_i is the ground-truth label for the i-th sample, $\hat{y}_i$ is the predicted label for the i-th sample, and $\mathbb{I}(\cdot)$ is the indicator function that equals 1 if its argument is true, and 0 otherwise.

Loss Function: We adopt the standard cross-entropy loss for multi-class classification. Given model output logits $\mathbf{z} \in \mathbb{R}^{B\times C}$ and ground truth labels $y \in \{1,\ldots,C\}^B$, the loss is computed as:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{B}\sum_{i=1}^{B}\log\left(\frac{e^{z_{i,y_i}}}{\sum_{c=1}^{C} e^{z_{i,c}}}\right) \tag{21}$$

where $z_{i,c}$ denotes the predicted score of the i-th sample for class c, y_i is the ground-truth label for the i-th sample, C is the number of classes, and B is the batch size.

FLOPs: We use FLOPs (Floating Point Operations) to quantify the computational complexity of each model.

Training Time: For efficiency analysis, we report the total training time measured across three repeated runs, each employing early stopping based on the criterion of no improvement in validation loss for five consecutive epochs.

statistical metrics: The Coefficient of Variation and 95% Confidence Intervals were considered for all models across three datasets. Overall, SWS-Net exhibits the most stable training time and accuracy, while other models show larger variations in certain datasets.

Implementation Details

For the CSI data, we train all models using the Adam optimizer with a learning rate of 0.01, a batch size of 16, and 32 channels, for up to 100 epochs. For the BVP data, we adopt the RMSprop optimizer with a learning rate of 0.001, a batch size of 32, and configure the network with 32 channels and a GRU hidden size of 196. Early stopping is applied in both cases if the validation loss does not improve for 5 consecutive epochs. We used the default membrane threshold (1.0) without any tuning. Since each time step aligns with the BVP and CSI signal sampling rate, the network's temporal integration is naturally consistent with the signal dynamics. The final performance is reported as the average over three independent runs. All experiments are implemented in PyTorch and conducted on 1–2 NVIDIA A100-SXM4-40GB GPUs.

Model	NTU-Fi-HumanID				NTU-Fi-HAR				Widar3.0 (22 gestures)			
	F(G)	P(M)	T(s)	TR	F(G)	P(M)	T(s)	TR	F(G)	P(M)	T(s)	TR
CNN	42.09	1.91	173.95	1.17×	42.09	1.91	145.84	1.13×	25.03	0.38	5810.18	12.96×
CNNRes	62.49	1.75	172.53	1.16×	62.49	1.75	140.94	1.09×	11.22	0.52	1341.23	2.99×
LSTM	18.56	2.28	327.42	2.20×	18.56	2.28	168.89	1.31×	1.84	0.46	494.09	1.10×
Transformer	43.34	1.62	191.11	1.29×	43.34	1.62	139.62	1.08×	1.27	0.54	996.53	2.22×
CNNGRU	–	–	–	–	–	–	–	–	0.45	0.50	983.80	2.19×
SWS-Net	**4.62**	**1.60**	**148.73**	1.00×	**4.62**	**1.60**	**128.80**	1.00×	**0.24**	**0.36**	**448.29**	1.00×

Table 1: Comparison of Model FLOPs (F) and Params (P), Training Time (T) and Time Ratio (TR)

Model	NTU-Fi-HumanID			NTU-Fi-HAR			Widar3.0 (22 gestures)		
	Acc (%)	P (M)	PR	Acc (%)	P (M)	PR	Acc (%)	P (M)	PR
CNN	97.01	1.908	1.19×	**100**	1.908	1.19×	70.73	0.376	1.03×
CNNRes	97.86	1.752	1.09×	99.75	1.752	1.09×	69.84	0.517	1.42×
LSTM	95.85	2.278	1.42×	99.62	2.278	1.42×	80.17	0.463	1.27×
Transformer	98.78	1.624	1.01×	98.99	1.624	1.01×	82.98	0.542	1.49×
CNNGRU	–	–	–	–	–	–	80.87	0.495	1.36×
ViT (Yang et al. 2023)	76.84	1.054	0.66×	93.75	1.052	0.66×	67.72	**0.106**	0.29×
SenseMamba (Huang et al. 2025)	99.34	**0.021**	0.01×	99.38	**0.021**	0.01×	66.15	0.194	0.29×
ProbSparse Attention (Yi et al. 2024)	**99.93**	3.215	2.00×	**100**	3.215	2.00×	–	–	–
WiGaiNet (Wang et al. 2025a)	99.32	7.3	4.55×	–	–	–	–	–	–
LiteWiHAR (Liu et al. 2024)	99.83	0.45	0.28×	–	–	–	–	–	–
VBCNet (Ge et al. 2024)	–	–	–	98.49	–	–	77.92	–	–
CaiT (Luo et al. 2024)	–	–	–	99.2	2.3	1.43×	–	–	–
LWiHS (Liu et al. 2025)	–	–	–	99.9	2.75	1.71×	–	–	–
RGANet (Hu et al. 2025)	–	–	–	97.95	0.44	0.27×	–	–	–
SWS-Net	98.11	1.604	1.00×	**100**	1.604	1.00×	**83.83**	0.364	1.00×

Table 2: Comparison of Accuracy, Parameters (P) and Parameter Ratio (PR) across Three Datasets

Experimental Results

To evaluate the training dynamics and generalization ability of different models, we first visualize the validation loss curves across all tasks during 25 epochs, as shown in Figure 2. The experiments are conducted on three representative WiFi sensing tasks: NTU-Fi-HumanID, NTU-Fi-HAR and Widar3.0 22-gestures.

The SWS-Net consistently shows the fastest convergence and lowest validation loss across all three datasets. On NTU-Fi-HumanID (Figure 2a), it quickly reduces the loss in the first few epochs and stabilizes below 0.3, outperforming all baselines. For NTU-Fi-HAR (Figure 2b), while other methods show early fluctuations, our model decreases smoothly and reaches near-zero loss by epoch 10. On the more challenging Widar3.0 22-gestures dataset (Figure 2c), it again achieves the best convergence and the lowest final loss.

We compare the computational complexity and training efficiency of each model across different datasets, as summarized in Table 1. The table presents the number of floating-point operations (FLOPs), model size (in MB), total training time per dataset, and the average time ratio — defined as the ratio between the average training time of each baseline and that of our SWS-Net. A higher ratio indicates longer training times relative to our model. As shown, SWS-Net demonstrates superior efficiency across all datasets. On NTU-Fi, it requires only 4.62G FLOPs and 1.60M parameters, while on Widar3.0, it uses merely 0.24G FLOPs and 0.364M parameters. The model also trains 1.08 to 12.96 times faster than competing approaches.

Table 2 presents a comprehensive comparison of classification accuracy, model size (in millions of parameters), and the normalized parameter ratio across three datasets. On the Widar3.0 (22 gestures) dataset, our model achieves the highest accuracy of 83.83%, outperforming all baseline methods. It maintains a compact parameter size of only 0.36M, demonstrating its excellent efficiency. For the NTU-Fi-HumanID dataset, our model attains 98.11% accuracy, closely following the best-performing model (ProbSparse Attention, 99.93%) but with only half the number of parameters (1.604M vs. 3.215M), and significantly fewer than large models like WiGaiNet (7.3M). On NTU-Fi-HAR, our model reaches 100% accuracy, matching the best baselines while using substantially fewer parameters than ProbSparse Attention (3.215M) and other large-scale architectures such as LWiHS (2.75M) and CaiT (2.3M). Compared to recent lightweight models like SenseMamba and ViT, which have fewer parameters, our method achieves much higher accuracy, especially on Widar3.0.

Conclusion

This work presents the integration of spiking neural networks into WiFi-based indoor sensing to enhance robustness under noisy conditions. A lightweight ANN-SNN hybrid architecture SWS-Net is proposed, with modality-specific preprocessing for CSI and BVP signals. Theoretical analysis is provided to compare the noise-handling capabilities of ANN and SNN models in WiFi scenarios. Experimental results across three benchmark datasets confirm the model's superior accuracy, faster convergence, and improved noise tolerance compared to conventional baselines. These findings demonstrate the potential of spike-based architectures for efficient and reliable indoor sensing.

Acknowledgments

This work was supported by the Guangdong Provincial Science and Technology Program (Grant No. 2024B0101010003), the National Natural Science Foundation of China No. 62306184, Natural Science Foundation of Top Talent of SZTU (Grant no. GDRC202320), Shenzhen Science and Technology Program (Grant no. JCYJ20240813113218025) and the National Key R&D Program of China (Grant No. 2024YFE0200801 and 2024YFE0200804).

References

Castagnetti, A.; Pegatoquet, A.; and Miramond, B. 2023. Neural information coding for efficient spike-based image denoising. *arXiv preprint*, 2305.11898.

Clancey, W. J. 1979. *Transfer of Rule-Based Expertise through a Tutorial Dialogue*. Ph.D. diss., Dept. of Computer Science, Stanford Univ., Stanford, Calif.

Clancey, W. J. 1983. Communication, Simulation, and Intelligent Agents: Implications of Personal Intelligent Machines for Medical Education. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI-83)*, 556–560. Menlo Park, Calif: IJCAI Organization.

Clancey, W. J. 1984. Classification Problem Solving. In *Proceedings of the Fourth National Conference on Artificial Intelligence*, 45–54. Menlo Park, Calif.: AAAI Press.

Clancey, W. J. 2021. The Engineering of Qualitative Models. Forthcoming.

Engelmore, R.; and Morgan, A., eds. 1986. *Blackboard Systems*. Reading, Mass.: Addison-Wesley.

Fan, N.; Tian, Z.; Dubey, A.; Deshmukh, S.; Murch, R.; and Chen, Q. 2024. Multitarget device-free localization via cross-domain Wi-Fi RSS training data and attentional prior fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 91–99.

Ge, F.; Dai, Z.; Yang, Z.; Wu, F.; and Tan, L. 2024. VBC-Net: A Hybird Network for Human Activity Recognition. *Sensors*, 24(23): 7793.

Guo, L.; Zhao, Q.; Wu, Y.; and Xu, G. 2022. Anti-interference of a small-world spiking neural network against pulse noise. *Applied Intelligence*, 52: 109645.

Guo, M.; Wang, Y.; Xu, G.; et al. 2023. Anti-Disturbance of Scale-Free Spiking Neural Network against Impulse Noise. *Brain Sciences*, 13(5): 837.

Hasling, D. W.; Clancey, W. J.; and Rennels, G. 1984. Strategic explanations for a diagnostic consultation system. *International Journal of Man-Machine Studies*, 20(1): 3–19.

He, Y.; Liu, J.; Li, M.; Yu, G.; and Han, J. 2024. Forward-Compatible Integrated Sensing and Communication for WiFi. *IEEE Journal on Selected Areas in Communications*.

He, Y.; Wei, M.; Li, D.; Li, P.; and Li, H. 2025. CFNet: CSI compression feedback network based on WiFi sensing. *Engineering Research Express*, 7(1): 015221.

Hu, J.; Ge, F.; Cao, X.; and Yang, Z. 2025. RGANet: A Human Activity Recognition Model for Extracting Temporal and Spatial Features from WiFi Channel State Information. *Sensors*, 25(3): 918.

Huang, Y.; Liu, J.; Shi, X.; Zhao, S.; Mi, T.; and Qiu, R. C. 2025. SenseMamba: A General Lightweight State Space Model for Wireless Human Sensing. *IEEE Sensors Journal*.

Li, B.; Cui, W.; Wang, W.; Zhang, L.; Chen, Z.; and Wu, M. 2021. Two-stream convolution augmented transformer for human activity recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 286–293.

Li, R.; Deng, T.; Feng, S.; Sun, M.; and Jia, J. 2025a. ConSense: Continually Sensing Human Activity with WiFi via Growing and Picking. In *Proceedings of the AAAI Conference on Artificial intelligence*, volume 39, 14292–14300.

Li, T.; Fan, L.; Yuan, Y.; and Katabi, D. 2022. Unsupervised learning for human sensing using radio signals. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3288–3297.

Li, Z.; Wu, Q.; Zhang, J.; Zhang, K.; and Wang, J. 2025b. Noise-Injected Spiking Graph Convolution for Energy-Efficient 3D Point Cloud Denoising. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(17): 18629–18637.

Liang, X.; Yang, H. H.; Guo, K.; and Quek, T. Q. 2024. Enabling Respiration Sensing via Commodity WiFi 6E Devices. In *GLOBECOM 2024-2024 IEEE Global Communications Conference*, 4382–4387. IEEE.

Liu, C.; Hao, Y.; Liu, Y.; Zhang, X.; Wang, X.; and Liu, Y. 2025. LWiHS: A Lightweight WiFi-Enabled Human Sensing Using Feature Fusion Strategy. *IEEE Transactions on Instrumentation and Measurement*.

Liu, C.; Liu, Y.; Hao, Y.; and Zhang, X. 2024. LiteWi-HAR: A Lightweight WiFi-Based Human Activity Recognition System. In *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, 1–5. IEEE.

Liu, S.; Zhao, Y.; and Chen, B. 2017. WiCount: A deep learning approach for crowd counting using WiFi signals. In *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, 967–974. IEEE.

Luo, F.; Khan, S.; Jiang, B.; and Wu, K. 2024. Vision Transformers for Human Activity Recognition Using WiFi Channel State Information. *IEEE Internet of Things Journal*, 11(17): 28111–28122.

Moshiri, P. F.; Shahbazian, R.; Nabati, M.; and Ghorashi, S. A. 2021. A CSI-based human activity recognition using deep learning. *Sensors*, 21(21): 7225.

NASA. 2015. Pluto: The 'Other' Red Planet. https://www.nasa.gov/nh/pluto-the-other-red-planet. Accessed: 2018-12-06.

Rice, J. 1986. Poligon: A System for Parallel Problem Solving. Technical Report KSL-86-19, Dept. of Computer Science, Stanford Univ.

Robinson, A. L. 1980. New Ways to Make Microcircuits Smaller. *Science*, 208(4447): 1019–1022.

Shi, J.; Zhang, B.; Dubey, A.; Murch, R.; and Jing, L. 2024. Vision Reimagined: AI-Powered Breakthroughs in WiFi Indoor Imaging. *arXiv preprint arXiv:2401.04317*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention Is All You Need. arXiv:1706.03762.

Wang, C.; Fu, X.; Yang, Z.; and Li, S. 2025a. NeuralWiGait: An Accurate WiFi-Based Gait Recognition System Using Hybrid Deep Learning Framework. *The Journal of Supercomputing*, 81(2): 373.

Wang, D.; Yang, J.; Cui, W.; Xie, L.; and Sun, S. 2022. CAUTION: A Robust WiFi-based Human Authentication System via Few-shot Open-set Gait Recognition. *IEEE Internet of Things Journal*.

Wang, H.; Li, X.; Li, J.; Zhu, H.; and Luo, J. 2025b. VR-Fi: Positioning and Recognizing Hand Gestures via VR-embedded Wi-Fi Sensing. *IEEE Transactions on Mobile Computing*.

Wei, Z.; Chen, W.; Ning, S.; Lin, W.; Li, N.; Lian, B.; Sun, X.; and Zhao, J. 2025. A Survey on WiFi-based Human Identification: Scenarios, Challenges, and Current Solutions. *ACM Transactions on Sensor Networks*, 21(1): 1–32.

Yang, J.; Chen, X.; Zou, H.; Lu, C. X.; Wang, D.; Sun, S.; and Xie, L. 2023. SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing. *Patterns*, 4(3).

Yang, J.; Chen, X.; Zou, H.; Wang, D.; Xu, Q.; and Xie, L. 2022. EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression. *IEEE Internet of Things Journal*, 9(15): 13086–13095.

Yang, J.; Zou, H.; Jiang, H.; and Xie, L. 2018. CareFi: Sedentary behavior monitoring system via commodity WiFi infrastructures. *IEEE Transactions on Vehicular Technology*, 67(8): 7620–7629.

Yi, D.; Zhang, H.; Feng, S.; Fang, J.; and Wang, W. 2024. ProbSparse attention with stacked group convolution for wireless signal-based human activity recognition. In *2024 16th WCSP*, 1349–1354. IEEE.

Yue, S.; He, H.; Wang, H.; Rahul, H.; and Katabi, D. 2018. Extracting multi-person respiration from entangled RF signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2): 1–22.

Zeng, Y.; Wu, D.; Xiong, J.; Liu, J.; Liu, Z.; and Zhang, D. 2020. MultiSense: Enabling multi-person respiration sensing with commodity WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3): 1–29.

Zhang, Y.; Zheng, Y.; Qian, K.; Zhang, G.; Liu, Y.; Wu, C.; and Yang, Z. 2021. Widar3. 0: Zero-effort cross-domain gesture recognition with wi-fi. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhang, Z.; Wang, J.; Xia, M.; Shi, C.; Wen, W.; and Zhang, D. 2025. Self-Attention-Enhanced PSW-LSTM for 3D Indoor Pedestrian Positioning with Integrated WiFi, Magnetometer and Barometer Sensors. *IEEE Transactions on Instrumentation and Measurement*.

Zhao, Z.; Chen, T.; Cai, Z.; Li, X.; Li, H.; Chen, Q.; and Zhu, G. 2025. Crossfi: A cross domain wi-fi sensing framework based on siamese network. *IEEE Internet of Things Journal*.

Zheng, Y.; Zhang, Y.; Qian, K.; Zhang, G.; Liu, Y.; Wu, C.; and Yang, Z. 2019. Zero-effort cross-domain gesture recognition with Wi-Fi. In *Proceedings of the 17th annual international conference on mobile systems, applications, and services*, 313–325.

Zou, H.; Jiang, H.; Yang, J.; Xie, L.; and Spanos, C. 2017a. Non-intrusive occupancy sensing in commercial buildings. *Energy and Buildings*, 154: 633–643.

Zou, H.; Yang, J.; Prasanna Das, H.; Liu, H.; Zhou, Y.; and Spanos, C. J. 2019. WiFi and vision multimodal learning for accurate and robust device-free human activity recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 0–0.

Zou, H.; Zhou, Y.; Yang, J.; Gu, W.; Xie, L.; and Spanos, C. 2017b. Freecount: Device-free crowd counting with commodity wifi. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, 1–6. IEEE.

Zou, H.; Zhou, Y.; Yang, J.; Gu, W.; Xie, L.; and Spanos, C. 2018. Wifi-based human identification via convex tensor shapelet learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.