Wuhan University
Journal of Natural Sciences

2026, Vol.31 No.1, 010-024

Article ID 1007-1202(2026)01-0010-15 DOI https://doi.org/10.1051/wujns/2026311010

Cite this article: HU Biling, TONG Yu. Human Activity Recognition Using a CNN with an Enhanced Convolutional Block Attention Module [J]. *Wuhan Univ J of Nat Sci*, 2026, 31(1): 10-24.

Human Activity Recognition Using a CNN with an Enhanced Convolutional Block Attention Module

□ **HU Biling[1,2], TONG Yu[1]**

1. School of Computer and Artificial Intelligence, Hefei Normal University, Hefei 236032, Anhui, China;

2. Anhui Engineering Laboratory for Sports Health Information Monitoring Technology (AEL—SHIMT) , Hefei 236032, Anhui, China

Abstract: WiFi-based human activity recognition (HAR) provides a non-intrusive approach for ubiquitous monitoring; however, achieving both high accuracy and robustness simultaneously remains a significant challenge. This paper proposes a Convolutional Neural Network with Enhanced Convolutional Block Attention Module (CNN-ECBAM) framework. The approach systematically converts raw Channel State Information (CSI) into pseudo-color images, effectively preserving essential signal characteristics for deep neural network processing. The core innovation is an Enhanced Convolutional Block Attention Module (ECBAM), tailored to CSI data characteristics, which integrates Efficient Channel Attention (ECA) and Multi-Scale Spatial Attention (MSSA). By employing learnable adaptive fusion weights, it achieves dynamic synergy between channel and spatial features, enabling the network to capture highly discriminative spatiotemporal patterns. The ECBAM module is integrated into a unified Convolutional Neural Network (CNN) to form the overall CNN-ECBAM model. Experimental results on the UT-HAR and NTU-Fi_HAR datasets demonstrate that CNN-ECBAM achieves competitive performance in recognition accuracy and outperforms mainstream baseline models. Specifically, it attains 99.20% accuracy on UT-HAR (surpassing ResNet-18 at 98.60%) and achieves 100% accuracy on NTU-Fi_HAR (exceeding GAF-CNN at 99.62%). These results validate the effectiveness of the proposed method for high-precision and reliable WiFi-based HAR.

Key words: human activity recognition; deep learning; channel state information; Enhanced Convolutional Block Attention Module (ECBAM); pseudo-color images

CLC number: TP391.4

0 Introduction

Ubiquitous sensing technology is a promising research frontier in the IoT sensing domain and has broad application prospects in areas such as security, sports monitoring, health monitoring, and entertainment[1]. Within this landscape, human activity recognition (HAR) is one of the key research topic[2-3]. Traditional HAR relies on vision-based systems or wearable sensors, which face limitations: vision requires line-of-sight and raises privacy concerns, while wearables impose user compliance challenges. In contrast, ubiquitous WiFi infrastructure enables device-free, non-intrusive activity sensing through wireless signal analysis. In particular, Channel State Information (CSI) in WiFi signals has shown great potential in a variety of device-free human

Received date: 2025-07-25

Foundation item: Supported by Anhui Provincial Engineering Research Center for Sports and Health Information Monitoring Technology(KF2023012)

Biography: HU Biling, female, Lecturer, research directions: wireless sensing, evolutionary computation. E-mail: bilinghu@hfnu.edu.cn

perception tasks, such as activity recognition[4-5], fall detection[6], gesture recognition[7], human identity recognition[8-9], and human counting[10]. WiFi CSI-based HAR offers distinct advantages: device-free operation, non-line-of-sight capability, privacy preservation, illumination independence, and cost-effectiveness.

Currently, there are two main types of WiFi-based sensing methods: model-based and learning-based. Model-based methods, representing early phases of development, primarily rely on physical signal propagation models. These evolved from early approaches using Received Signal Strength Indicator (RSSI)[11-12] to more precise methodologies leveraging CSI[13-16]. However, the coarse granularity of RSSI and the complexity of physical modeling for intricate human movements limit their scalability.

Conversely, learning-based methods leverage deep neural networks to extract features directly from data, achieving superior performance in complex sensing tasks. Despite their success, current deep learning approaches for HAR still face significant limitations: 1) LSTM-based models[17] that effectively capture temporal dependencies but neglect spatial relationships and joint spatial-temporal features, limiting discriminative capability; 2) CNN (Convolutional Neural Network) -based methods[18-19] that extract spatial features from CSI-transformed images yet suffer from shallow architectures lacking adaptive multi-scale attention; 3) CNN-LSTM hybrids[20-22] that jointly model spatial-temporal features but lack explicit multi-scale attention and adaptive fusion; 4) Domain-specific attention mechanisms[23-24] that improve robustness through temporal or separable dimensional attention but miss multi-scale spatial receptive fields; 5) Transformer-based approaches[25-27] that capture long-range dependencies at the cost of high computational complexity and environmental sensitivity. To address these limitations, we draw inspiration from the transformative success of attention mechanisms in NLP (Natural Language Processing)[28-30], computer vision[31-34], and time-series analysis[35-39]. While attention has shown emerging potential in WiFi sensing[40-42], standard mechanisms often fail to address the specific domain challenges of CSI data, such as efficiency, adaptability to signal variability, and the need for multi-scale spatial receptive fields[43-44]. Therefore, beyond merely applying existing attention models, there is a critical need for a domain-specific innovation. This motivates our development of an Enhanced CBAM framework, explicitly tailored to overcome both the inherent limitations of prior HAR methods and the specific shortcomings of conventional attention mechanisms in robust and efficient CSI-based recognition.

In this paper, we propose a novel CNN-ECBAM approach, a learning-based framework specifically designed to overcome the aforementioned challenges. By integrating an Enhanced Convolutional Block Attention Module (ECBAM) into a convolutional neural network, our method effectively captures discriminative spatio-temporal patterns from CSI data. The proposed approach consists of three main components: CSI-to-pseudo-color image conversion, Enhanced CBAM architecture, and integrated CNN framework.The main contributions of this paper are as follows:

1) Novel CSI-to-image Conversion Methodology: A systematic approach is introduced for converting raw CSI signals into pseudo-color images that preserve essential signal characteristics while enabling effective CNN processing. The method incorporates optimized normalization, color mapping, and resolution adaptation techniques specifically designed for WiFi CSI data.

2) ECBAM Architecture for CSI Data: An Enhanced Convolutional Block Attention Module is developed for pseudo-color images derived from WiFi CSI signals. The design integrates standard deviation pooling in channel attention and dilated convolutions in spatial attention, facilitating the capture of both signal variability and multiscale spatial patterns.

3) Comprehensive CNN-ECBAM Framework: A unified CNN architecture is constructed by integrating the ECBAM modules, forming the CNN-ECBAM framework. This design leverages the spatial feature extraction capability of CNNs together with the refined attention mechanisms of ECBAM to achieve superior activity recognition performance.

The remainder of this paper is organized as follows. Section 1 briefly introduces CSI, CBAM, and CNN. Section 2 presents the detailed methodology of our CNN-ECBAM approach, including the CSI-to-pseudo-color image conversion process, ECBAM architecture design, and integrated CNN framework. Section 3 describes the experimental setup, datasets, evaluation metrics, training strategy used in our study. Section 4 presents the experimental results and comprehensive analysis. Section 5 concludes with a summary of contributions and practical implications.

1 Preliminaries and Background

1.1 Channel State Information (CSI)

Wireless signal propagation is sensitive to environmental factors, particularly human motion, which induces reflection, refraction, scattering, and diffraction. In a fixed environment, distinct movements produce characteristic temporal fluctuations in the received channel response. By learning the statistical mapping between these fluctuations and the underlying motion states, human activities can be inferred from variations in WiFi Channel State Information (CSI). Figure 1 presents the overall framework of WiFi CSI-based activity recognition, comprising signal acquisition, processing, and classification stages.

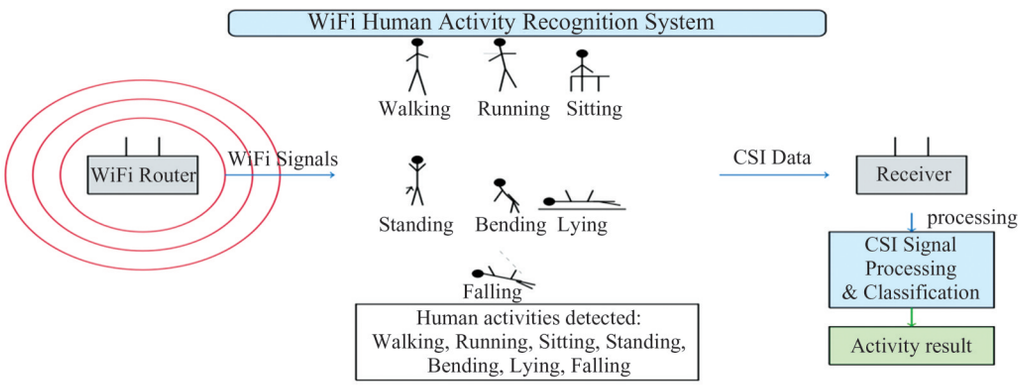


Fig.1 The overview of human activity recognition

The wireless channel can be expressed in the frequency domain as follows:

$$\boldsymbol{Y}=\boldsymbol{H}\cdot\boldsymbol{X}+\boldsymbol{N}, \tag{1}$$

where $\boldsymbol{H}$ is the channel state information matrix, $\boldsymbol{Y}$ is the received signal vector, $\boldsymbol{X}$ is the transmitted signal vector, $\cdot$ denotes matrix multiplication, and $\boldsymbol{N}$ is the additive Gaussian white noise vector. $\boldsymbol{H}$ can be further expressed as:

$$\boldsymbol{H}(j)=|H(j)|\mathrm{e}^{\mathrm{i}\sin\angle H(j)}. \tag{2}$$

Here $H(j)$ denotes the channel state information of the j-th subcarrier, which contains two components, CSI magnitude ($|H(j)|$) and phase ($\angle H(j)$). According to multiple-input multiple-output (MIMO) technology, the combined CSI of the data streams obtained by combining multiple antennas at the transceiver end with each other can be expressed with matrix form as:

$$\boldsymbol{H}=\begin{bmatrix} H_{11} & \cdots & H_{1m} \\ \vdots & H_{ij} & \vdots \\ H_{n1} & \cdots & H_{nm} \end{bmatrix}, \tag{3}$$

where H_{ij} denotes the CSI between the i-th receiving antenna and the j-th transmitting antenna, and m and n denote the number of antennas at the receiving and transmitting ends, respectively.

1.2 Convolutional Block Attention Module (CBAM)

CBAM[34] enhances CNN feature representations through cascaded channel and spatial attention mechanisms, as shown in Fig. 2. Channel attention weights feature maps based on inter-channel relationships, while spatial attention emphasizes informative regions within each feature map, jointly improving network discriminability.

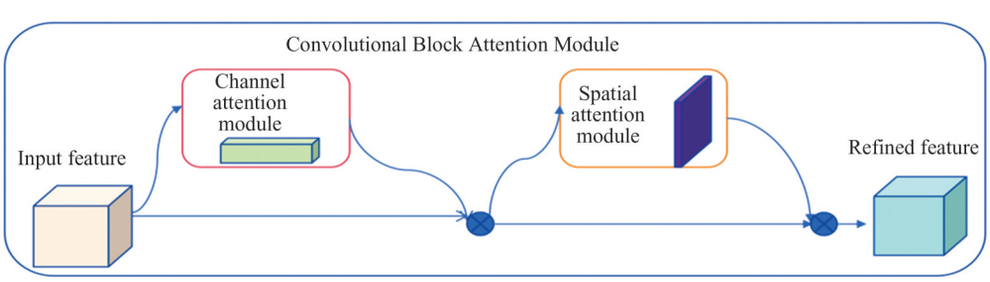


Fig. 2 The overview of CBAM

Given an original feature map $F \in \mathbb{R}^{C\times H\times W}$ as input, CBAM sequentially derives a one-dimensional channel attention map $M_c \in \mathbb{R}^{C\times 1\times 1}$ and a two-dimensional spatial attention map $M_s \in \mathbb{R}^{1\times H\times W}$. The one-dimensional channel attention map M_c will be multiplied channel by channel with the original feature map to obtain the channel-weighted feature map F', F' will be multiplied position by position with the two-dimensional spatial attention map M_s to obtain the final output. The overall attention process can be summarized as

$$\begin{aligned} F' &= M_c(F)\otimes F, \\ F'' &= M_s(F')\otimes F', \end{aligned} \tag{4}$$

where $\otimes$ denotes element-by-element multiplication.

For the Channel Attention Module (CAM) in Fig.2, its goal is to learn the importance weights of each channel to enhance useful channels and suppress redundant channels. Its realization steps include three phases: feature compression, weight generation and feature re-calibration. Feature compression performs Global Average Pooling (GAP) and Global Maximum Pooling (GMP) on the input feature map $F \in \mathbb{R}^{C\times H\times W}$ to obtain two $C\times 1\times 1$ channel description vectors, respectively. Weight generation takes the two vector inputs into the shared multilayer perceptron (MLP) and generates the channel attention weights $M_c \in \mathbb{R}^{C\times 1\times 1}$. The formula is described as follows:

$$M_c(F)=\sigma(\mathrm{MLP}(\mathrm{GAP}(F))+\mathrm{MLP}(\mathrm{GMP}(F))), \tag{5}$$

where σ is the Sigmoid function. The feature re-calibration multiplies M_c with the original feature map channel by channel to get the channel-weighted feature map. The spatial attention module learns the importance of each spatial location within feature maps, thereby focusing on key regions. Its implementation comprises four sequential stages: channel compression, feature concatenation, weight generation, and feature re-calibration. Channel compression performs average pooling and maximum pooling on the weighted feature map along the channel dimensions to obtain two $1\times \boldsymbol{H}\times \boldsymbol{W}$ feature maps. Feature splicing splices the two feature maps into $2\times \boldsymbol{H}\times \boldsymbol{W}$. Weight generation compresses the channels to 1 by a 7×7 convolutional layer to generate spatial attention weights $M_s \in \mathbb{R}^{1\times H\times W}$. The formula is as follows:

$$M_s(F)=\sigma(f^{7\times 7}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)])). \tag{6}$$

Feature re-calibration: Multiply M_s with the feature map position by position to get the final output.

1.3 Convolutional Neural Networks

The convolutional neural network was first introduced for handwritten ZIP code recognition [45] and was later formalized in their comprehensive work on document recognition[46]. These networks overcome the limitations of multilayer perceptrons (MLPs) via two key mechanisms: parameter-sharing convolutional filters and hierarchical spatial subsampling. The core computational framework of CNNs involves layered feature extraction through learnable convolution operators followed by non-linear downsampling. Mathematically, the discrete convolution operation applies a filter kernel to an input tensor through sliding window operations, generating feature maps via element-wise multiplication and summation. Spatial pooling layers subsequently reduce feature dimensionality by aggregating local regions through maximization (max-pooling) or averaging operations. A standard CNN architecture comprises alternating convolutional blocks, pooling layers, and fully connected classifiers, as illustrated in Fig. 3.

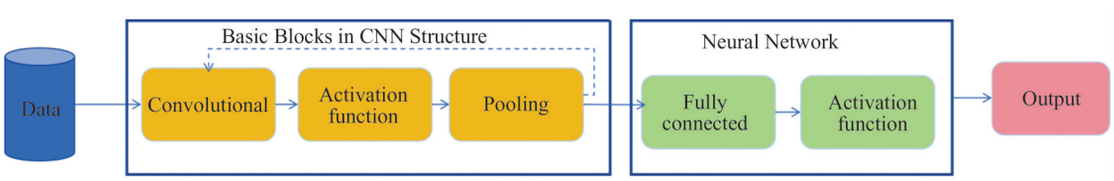


Fig.3 Main CNN architecture

2 Proposed Method

2.1 CSI-to-Pseudo-Color Image Preprocessing

The conversion of raw CSI signals into pseudo-color images involves several critical preprocessing steps designed to preserve essential signal characteristics while enabling effective CNN processing. This section provides detailed specifications of each preprocessing component.

2.1.1 Data normalization and scaling

The raw CSI amplitude data shows considerable variability across environments, devices, and time peri-

ods. To enhance training consistency and stability, a comprehensive normalization strategy is employed.

1) Global Min-Max Normalization: For the entire dataset $\mathcal{D}=\{X_1, X_2, \cdots, X_N\}$, where each X_i represents a CSI sample, we compute:

$$X_{\text{norm}} = \frac{X - \min(\mathcal{D})}{\max(\mathcal{D}) - \min(\mathcal{D})}, \tag{7}$$

where $\min(\mathcal{D})$ and $\max(\mathcal{D})$ are the global minimum and maximum values across all samples in the dataset.

2) Pixel Value Mapping: The normalized values are then mapped to the standard 8-bit pixel intensity range:

$$P = \text{round}(X_{\text{norm}} \times 255), \tag{8}$$

where $P \in [0, 255]$ represents the pixel intensity values, and the data type is converted to uint8 for memory efficiency and compatibility with image processing frameworks.

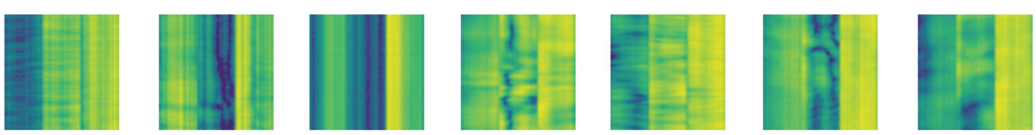
2.1.2 Color mapping and visualization

To convert single-channel normalized CSI data into informative pseudo-color representations, the viridis colormap is utilized. This colormap is widely adopted owing to its desirable properties: (i) perceptual uniformity, ensuring that equal data variations correspond to equal perceptual differences; (ii) monotonic luminance, providing a consistent brightness progression with increasing values; and (iii) accessibility for individuals with color vision deficiencies, while maintaining high contrast across the entire dynamic range. Formally, the mapping function translates normalized intensity values into RGB triplets, thereby enhancing the interpretability of CSI visualizations.

2.1.3 Image standardization and signal quality validation

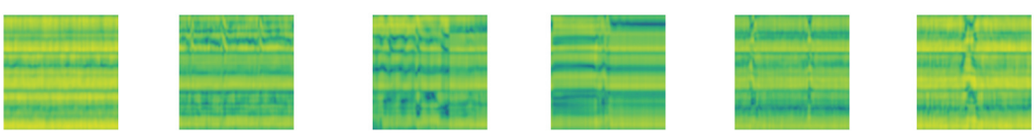
To ensure consistent input dimensions for the CNN architecture, pseudo-color images are generated under standardized settings. Each image is initially rendered at 8×8 inches with 100 DPI and subsequently rescaled to 64×64 pixels for computational efficiency. The images are represented in three RGB channels and stored as float32 tensors. To preserve the intrinsic structure of CSI data, the antenna and subcarrier dimensions are mapped to the vertical and horizontal axes, respectively, while temporal variations are encoded through color intensity.

Signal integrity during preprocessing is ensured through validation of dynamic range preservation, noise floor distinguishability, and retention of activity-specific signatures. This pipeline effectively transforms raw CSI measurements into discriminative pseudo-color images, which are then used as input to the CNN-ECBAM model. Representative samples are presented in Fig. 4 and Fig. 5.



(a) Bend (b) Fall (c) Lie down (d) Run (e) Sit down (f) Stand up (g) Walk

Fig. 4 UT_HAR generated RGB images



(a) Box (b) Circle (c) Clean (d) Fall (e) Run (f) Walk

Fig.5 NTU-Fi_HAR generated RGB images

2.2 Enhanced CBAM

The core contribution of our work lies in the Enhanced Convolutional Block Attention Module (ECBAM), which significantly improves upon the original CBAM by incorporating several innovative mechanisms specifically designed for pseudo-color images derived from WiFi CSI data. Our Enhanced CBAM addresses the unique characteristics of CSI signals by integrating four key innovations: (1) an Efficient Channel Attention (ECA) mechanism, (2) Multi-scale Spatial Attention with parallel convolutions, (3) an Adaptive Weight Learning mechanism for dynamic feature fusion, and (4) explicit residual connections for gradient flow optimization and feature preservation.Its detailed structure is shown in Fig. 6.

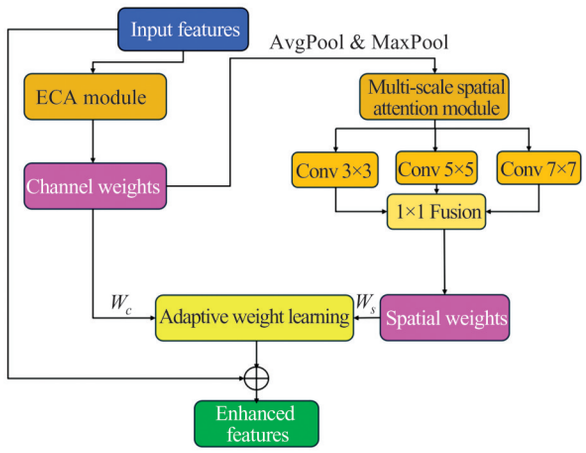


Fig.6 Enhanced CBAM architecture

2.2.1 Efficient Channel Attention (ECA) module

Traditional CBAM applies global average and max pooling for channel attention, but this design is limited in capturing the complex statistics of CSI-based pseudo-color images. In the proposed framework, the conventional module is replaced with Efficient Channel Attention (ECA), which achieves more effective channel-wise recalibration while preserving computational efficiency. ECA adaptively determines kernel sizes according to channel dimensions, thereby enhancing the modeling of cross-channel interactions. The adaptive kernel size is defined as:

$$t=|\log_2(c)+b|/\gamma, \tag{9}$$

$$k=\begin{cases} t, & \text{if } t \text{ is odd,} \\ t+1, & \text{others,} \end{cases} \tag{10}$$

where c is the number of input channels, $\gamma=2$ and $b=1$ are hyperparameters, and k is the final kernel size for the 1D convolution. Hyperparameters $\gamma=2$ and $b=1$ are adopted following the original ECA work[47], which determined these values through extensive experiments on large-scale datasets. The choice of $\gamma=2$ ensures logarithmic growth of the kernel sizes with respect to the channel dimensions, while $b=1$ provides an appropriate base offset for effective modeling of cross-channel interactions. The ECA mechanism then applies the 1D convolution along the channel dimension after global average pooling:

$$\mathrm{ECA}(F)=\sigma(\mathrm{Conv1D}_k(\mathrm{GAP(F)})), \tag{11}$$

where GAP denotes global average pooling, $\mathrm{Conv1D}_k$ represents 1D convolution with kernel size k, and σ is the sigmoid activation function.

The mathematical formulation for channel attention refinement is:

$$F'_c=F_c\otimes\mathrm{ECA}(F)_c, \tag{12}$$

where F_c represents the c-th channel of the input feature map F, and $\otimes$ denotes element-wise multiplication. $\mathrm{ECA}(F)_c$ is an importance weight computed for the c-th channel via global average pooling and 1D convolution. F'_c is the characterization of the channel after enhancement by the attentional mechanism, where important channels are reinforced and unimportant channels are suppressed.

2.2.2 Multi-scale spatial attention module

The spatial attention component of our ECBAM incorporates parallel convolutions with multiple kernel sizes to capture spatial patterns at different receptive fields. This design is particularly crucial for CSI-derived pseudo-color images, where spatial relationships manifest at various scales. The input to this module is generated by concatenating the channel-wise average-pooled and max-pooled features from the input tensor. This combined feature descriptor is then processed through three parallel convolution branches with 3×3, 5×5, and 7×7 kernels:

$$\mathrm{SA}_{3\times3}(F)=\mathrm{Conv}_{3\times3}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)]), \tag{13}$$

$$\mathrm{SA}_{5\times5}(F)=\mathrm{Conv}_{5\times5}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)]), \tag{14}$$

$$\mathrm{SA}_{7\times7}(F)=\mathrm{Conv}_{7\times7}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)]), \tag{15}$$

where $[\bullet;\bullet]$ denotes concatenation along the channel dimension, and the input to each branch is formed by concatenating channel-wise average and max pooled features.

The multi-scale features are then fused through a 1×1 convolution layer:

$$\mathrm{SA}_{\mathrm{multi}}(F)=\sigma(\mathrm{Conv}_{1\times1}([\mathrm{SA}_{3\times3}(F);\mathrm{SA}_{5\times5}(F);\mathrm{SA}_{7\times7}(F)])). \tag{16}$$

The final spatial attention map is generated using a sigmoid activation and applied to the feature map:

$$F''_{\mathrm{spatial}}=F'\otimes\mathrm{SA}_{\mathrm{multi}}(F'). \tag{17}$$

F''_{spatial} is a feature map enhanced by a spatial attention mechanism, where important spatial regions are reinforced and unimportant regions are suppressed. F' is the channel-enhanced feature map with c channels, serving as input to the spatial attention module. $\mathrm{SA}_{\mathrm{multi}}(F')$ is the attention weight map generated by the multi-scale spatial attention module, indicating the importance of spatial locations.

2.2.3 Adaptive weight learning mechanism

A critical innovation in our Enhanced CBAM is the introduction of learnable adaptive weights that dynamically balance the outputs of the attention pathways. This mechanism combines the channel-attended features with

the final spatially-attended features. The adaptive weight learning mechanism is formulated as:

$$W_{\text{channel}}, W_{\text{spatial}} \in \mathbb{R} > 0, \quad W_{\text{channel}}(0) = W_{\text{spatial}}(0) = 1.0, \tag{18}$$

where the initialization values are set to 1.0. The normalized weights are computed via softmax normalization:

$$w_c = \frac{\exp(W_{\text{channel}})}{\exp(W_{\text{channel}}) + \exp(W_{\text{spatial}})},$$

$$w_s = \frac{\exp(W_{\text{spatial}})}{\exp(W_{\text{channel}}) + \exp(W_{\text{spatial}})}, \tag{19}$$

and the final enhanced attention output is

$$F_{\text{out}} = w_c F' + w_s F''_{\text{spatial}}. \tag{20}$$

2.2.4 Residual connection integration

To maintain feature preservation and facilitate gradient flow, our Enhanced CBAM is integrated within residual blocks. The residual connection ensures that the original feature information is preserved while allowing the attention mechanism to refine the features:

$$F_{\text{residual}} = F + \text{Enhanced\_CBAM}(F). \tag{21}$$

F_{residual} denotes the final output feature map obtained after enhanced CBAM processing with residual connections. In contrast, F represents the original input feature map of the Enhanced CBAM residual block, and Enhanced_CBAM(F) indicates the feature map generated by the Enhanced CBAM module, which integrates both channel and spatial attention.

The residual connection design in our Enhanced CBAM serves two critical functions: (1) gradient flow preservation during backpropagation, which prevents vanishing gradient problems in deep networks, and (2) feature identity preservation, ensuring that the original input features are maintained while allowing the attention mechanism to provide refinements. As depicted in Fig. 6, the skip connection pathway directly adds the input feature map to the attention-processed output, creating a robust learning pathway that combines both original and attention-enhanced features.

2.3 CNN-ECBAM Architecture

The proposed CNN-ECBAM architecture is systematically constructed to optimize pseudo-color image processing through strategically integrated attention mechanisms. The comprehensive framework comprises hierarchically organized components that facilitate multi-scale feature learning and adaptive attention allocation. Figure 7 illustrates the architecture.

1) Input Processing Module: Raw CSI amplitude data is converted into 64×64×3 RGB pseudo-color images, encoding spatial-temporal patterns in a format compatible with convolutional processing.

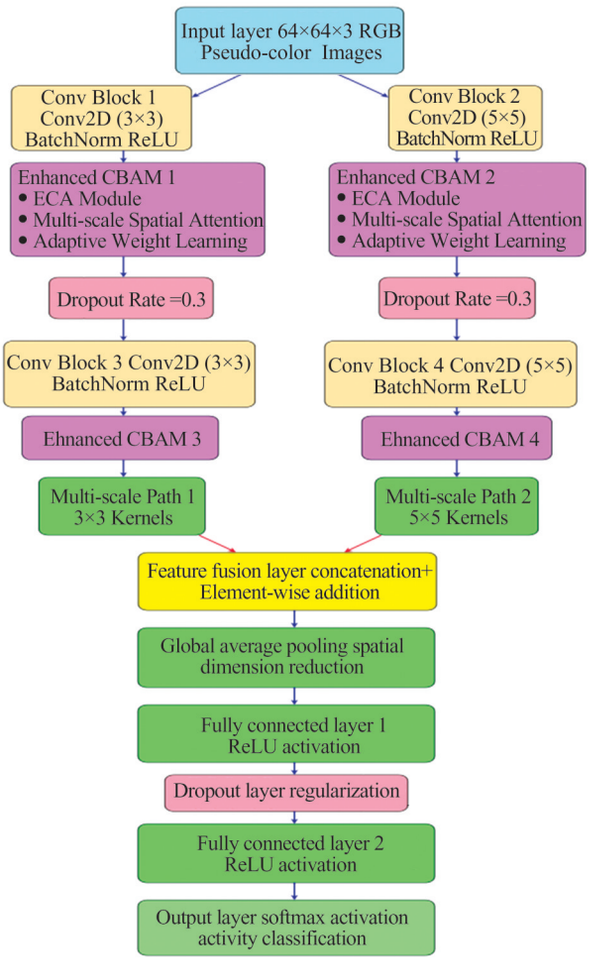


Fig.7 CNN-ECBAM architecture

2) Hierarchical Convolutional Framework: The network employs alternating 3×3 and 5×5 convolutions to capture local and contextual features, with batch normalization for training stability, ReLU activation, strategically positioned ECBAM modules for adaptive attention, and dropout (p=0.3) for regularization.

3) Multi-Resolution Feature Learning: Parallel convolutional branches with varying receptive fields simultaneously extract fine-grained local features and broader contextual patterns from CSI data.

4) Adaptive Feature Aggregation: Multi-scale features are integrated through concatenation (preserving feature diversity) and element-wise summation (dimension-preserving aggregation).

5) Decision Layer Architecture: Global average pooling reduces spatial dimensions while preserving channel information, followed by fully connected layers with ReLU activation and dropout, culminating in a softmax layer for activity classification.

3 Experimental Setup

3.1 Datasets

We evaluate our method on two public CSI-based HAR datasets: UT-HAR [17] and NTU-Fi_HAR [48]. UT-HAR comprises seven activity classes: bend, fall, lie down, run, sit down, stand up, and walk, collected using an Intel 5300 NIC with three antenna pairs, each capturing 30 subcarriers. Data collection was conducted in a controlled indoor environment.

NTU-Fi_HAR was acquired via Atheros CSI Tool and encompasses six activity classes: box, circle, clean, fall, run, and walk, representing diverse movement intensities and complexities.

3.2 Evaluation Metrics

To comprehensively assess the efficacy of our proposed methodology, we evaluate model performance using standard classification metrics: accuracy and F_1-score. Accuracy measures overall classification correctness, while F_1-score provides a balanced measure particularly valuable for imbalanced datasets. Finally, we present a comprehensive confusion matrix analysis to reveal inter-class confusion patterns.

3.3 Baseline Methods and Comparison

We compare CNN-ECBAM against the following baseline methods:

1) Classical Deep Learning Architectures: We evaluate against a modified LeNet-5 [46] where the original activation functions have been replaced with ReLU units to enhance training efficiency, and ResNet-18[49], which represents a standard convolutional neural network architecture widely adopted in computer vision tasks. We also incorporate the LSTM-based approach proposed in Ref. [17], which has shown effectiveness in capturing temporal dependencies for sequential CSI data in human activity recognition tasks.

2) WiFi-based HAR Specialized Methods: Our comparison includes ABLSTM [41], a prominent attention-based bidirectional LSTM approach specifically designed for WiFi sensing applications, and the CSI-to-image transformation method proposed in Ref. [19], which converts CSI data into Gramian Angular Field (GAF) representations followed by a four-layer CNN classifier for activity recognition.

To ensure the reliability of performance comparisons, paired t-tests are employed. Each model is trained and evaluated five times with different random seeds to derive performance distributions. Statistical significance is determined by comparing CNN-ECBAM against the strongest baseline on each dataset, where * indicates $p < 0.05$ and ** indicates $p < 0.01$.

3.4 Training Strategy

For all experiments, we train models for 200 epochs with adaptive batch size selection: LSTM models use batch size 64 for improved training stability, GAF-CNN follows the paper specification with batch size 32, while other models dynamically adjust batch sizes (32-200) based on GPU memory availability. The optimization strategy varies by model type: LSTM models on UT_HAR use Adam optimizer (lr = 0.001) with cosine annealing scheduling and gradient clipping (max_norm = 1.0), while LSTM models on NTU-Fi_HAR follow paper specifications using SGD optimizer (lr = 0.001, momentum = 0.9). GAF-CNN strictly adheres to paper requirements with Adam optimizer (lr = 0.001, weight_decay = 0.000 1), and other models use Adam with Reduce LR On Plateau scheduling. We implemented mixed-precision training for memory-intensive models (CNN-ECBAM, ResNet-18, GAF-CNN) to optimize GPU utilization, employ gradient accumulation for large datasets, and apply Xavier normal initialization for convolutional layers with forgetting gate bias set to 1.0 for LSTM models. Additionally, we incorporated early stopping mechanism with patience = 10, dropout regularization (0.3-0.5), and systematic memory management including periodic cleanup to ensure stable training across all experimental configurations.

4 Results and Analysis

4.1 Overall Performance Comparison

Table 1 summarizes the classification accuracy and F_1-scores for all models across both datasets. The results demonstrate the distinct robustness of our proposed CNN-ECBAM architecture. On the UT-HAR dataset, CNN-ECBAM achieved the highest accuracy of 99.20%, outperforming the strong ResNet-18 baseline (98.60%) and significantly surpassing ABLSTM (94.40%) and LeNet-5 (94.20%). Notably, GAF-CNN failed to capture discriminative features, yielding only 68.00% accuracy.

Conversely, on the NTU-Fi_HAR dataset, CNN-ECBAM attained 100% accuracy, confirming its superior feature extraction capability. A critical finding is the

performance volatility of baseline models: while GAF-CNN improved dramatically to 99.62% (indicating dataset-specific specialization), ResNet-18 degraded significantly to 87.50%, revealing severe overfitting. In contrast, CNN-ECBAM maintained state-of-the-art performance across both scenarios. Statistical significance testing further confirms these improvements: the gain over the strongest baselines is significant at $p < 0.05$ for UT-HAR and $p < 0.01$ for NTU-Fi_HAR, validating the effectiveness of the proposed attention-enhanced framework in overcoming the generalization limitations observed in conventional architectures.

Table 1 Performance comparison of different models on the UT-HAR and NTU-Fi_HAR datasets

%

Method	UT-HAR		NTU-Fi_HAR	
	Acc.	F_1	Acc.	F_1
Modified LeNet-5	94.20	94.17	97.73	97.72
ResNet-18	98.60	98.59	87.50	87.30
LSTM	87.00	86.77	95.83	95.79
ABLSTM	94.40	94.39	96.59	96.54
GAF-CNN	68.00	66.83	99.62	99.62
CNN-ECBAM	99.20*	99.20*	100**	100**

Note: Statistical significance vs. strongest baseline is indicated by * (p<0.05), ** (p<0.01) using paired t-test over 5 independent runs.

4.2 Convergence Analysis for Different Methods

This section provides a comparative analysis of convergence behavior, final accuracy, and generalization capability across all models on the UT-HAR dataset and NTU-Fi_HAR dataset , as illustrated in Fig. 8 and Fig. 9.

The proposed CNN-ECBAM demonstrated consistent stability and high accuracy across both datasets. On UT-HAR dataset, it converged to approximately 98% accuracy, matching the stability of ResNet-18 but with superior generalization. Notably, on the NTU-Fi_HAR dataset, where ResNet-18 exhibited severe overfitting (plateauing at 87% test accuracy despite 100% training accuracy), CNN-ECBAM maintained robust performance with minimal generalization gap. This contrast indicates that while standard deep architectures like ResNet-18 risk overfitting on specific CSI datasets, the domain-specific attention mechanisms in CNN-ECBAM effectively regularize the learning process. The Modified LeNet-5 achieved respectable stability but generally trailed the deeper architectures in peak accuracy.

A comparison between LSTM and ABLSTM highlights the critical role of attention mechanisms in temporal modeling. The standard LSTM struggled with volatility and lower accuracy (plateauing around 85% on UT-HAR), whereas ABLSTM achieved significantly more stable convergence (reaching 94% on UT-HAR and 97% on NTU-Fi_HAR). This confirms that attention mechanisms successfully mitigate the limitations of standard RNNs in capturing long-range dependencies within noisy CSI data.

The GAF-CNN model displayed significant performance variability. It failed to converge effectively on UT-HAR (fluctuating between 60%-70%) and exhibited extreme testing instability on NTU-Fi_HAR despite high training accuracy. This suggests that the GAF image representation may not consistently preserve the discriminative signal characteristics required for robust generalization across diverse activity sets.

In conclusion, while baseline models exhibited dataset-dependent strengths and weaknesses—often suffering from overfitting (ResNet-18) or instability (LSTM, GAF-CNN) —the CNN-ECBAM framework demonstrated better adaptability. By integrating efficient channel and multi-scale spatial attention, the model effectively balances model complexity with feature selection, ensuring robust convergence and generalization across diverse sensing environments.

4.3 Confusion Matrix for Different Methods

This section evaluates the inter-class discriminative capabilities of the models using confusion matrices for both datasets, as visualized in Fig. 10 and Fig. 11.

On the UT-HAR dataset, the proposed CNN-ECBAM demonstrated superior feature extraction, achieving near-perfect classification for distinct activities (fall, bend, run, sit down) and maintaining high pre-

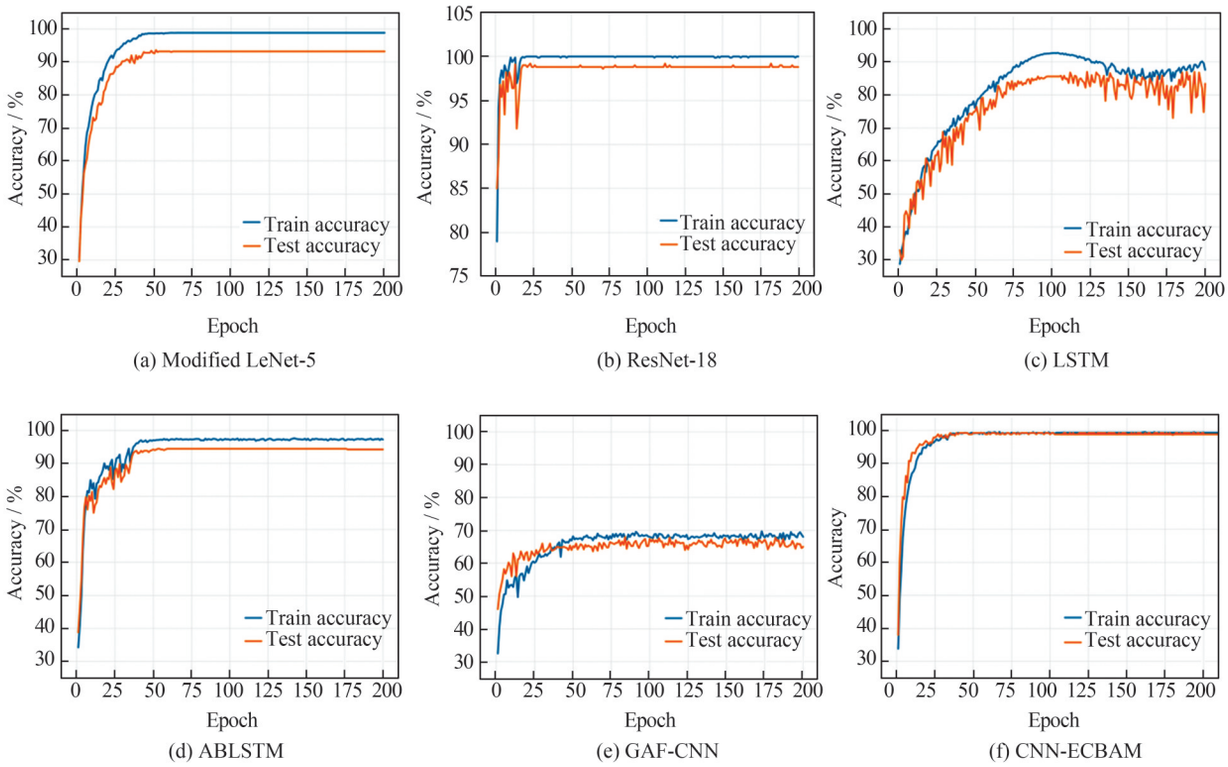


Fig.8 Accuracy convergence curves for all evaluated models on the UT_HAR dataset

Note: Each subplot shows the training accuracy (blue) and test accuracy (orange) over 200 epochs.

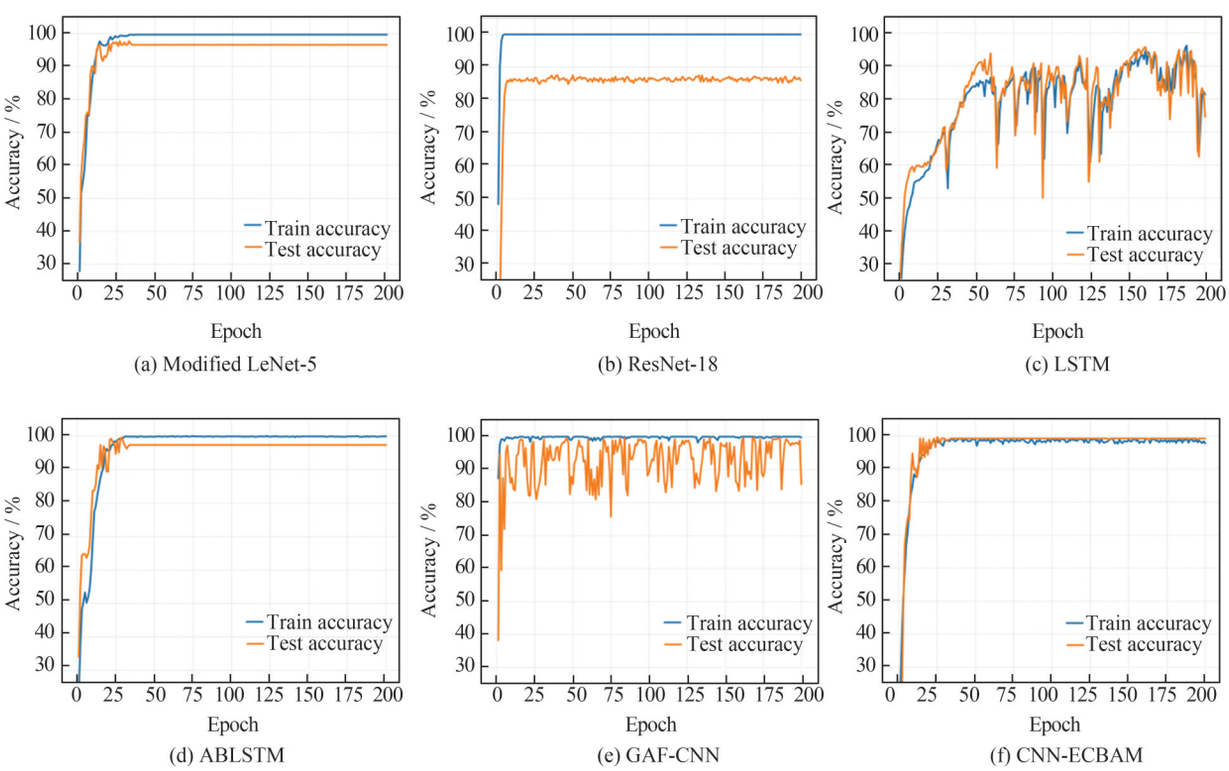


Fig.9 Accuracy convergence curves for all evaluated models on the NTU-Fi_HAR dataset

Note: Each subplot shows the training accuracy (blue) and test accuracy (orange) over 200 epochs.

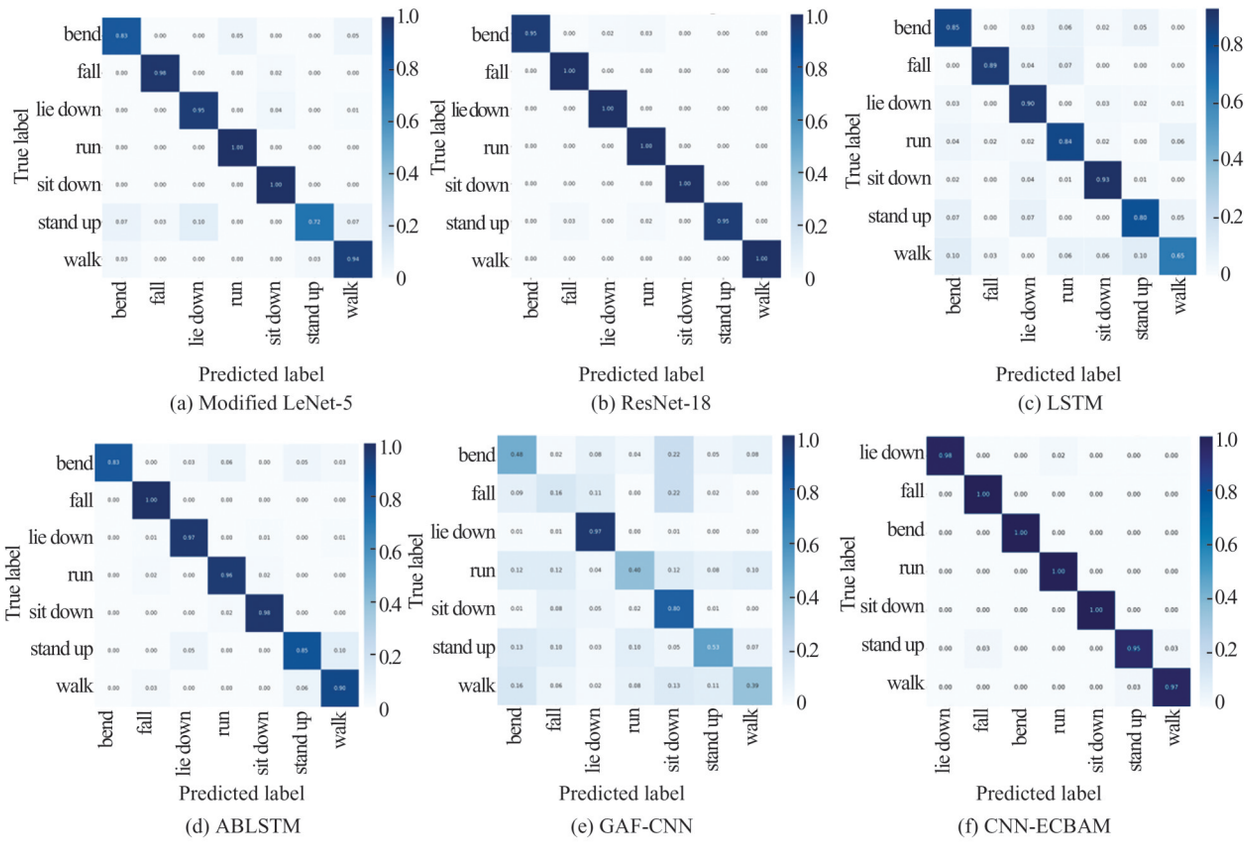


Fig.10 Confusion matrices for all evaluated models on the UT_HAR dataset

cision (>95%) even for challenging transitional actions (e. g., stand up vs. walk). ResNet-18 also performed strongly but exhibited minor confusion between specific postures like standup and lie down. In contrast, ABLSTM and modified LeNet-5 struggled more significantly with dynamic transitions, while LSTM and GAF-CNN showed widespread misclassification, indicating a failure to capture discriminative features for complex human activities.

Conversely, on the NTU-Fi_HAR dataset, CNN-ECBAM achieved perfect classification across all six activities (box, circle, clean, fall, run, walk). Notably, GAF-CNN, which performed poorly on UT-HAR, achieved high accuracy here, suggesting its effectiveness is highly dataset-dependent. However, ResNet-18 failed to generalize, showing significant confusion across multiple classes (e. g., misclassify walk as box or clean), which aligns with the overfitting observed in the convergence analysis.

Qualitative examination of misclassified samples reveals that errors predominantly occur at motion boundaries where activity transitions are ambiguous (e. g., the initiation of standup resembling walk). While CNN-ECBAM significantly reduces these errors through its multi-scale attention mechanism, the persistence of minor boundary confusions suggests that incorporating extended temporal context could further enhance performance in future work.

4.4 Ablation Studies

Systematic ablation studies were conducted across eight configurations to validate the contributions of individual Enhanced CBAM components: 1) CNN without attention, 2) original CBAM(baseline), 3) Efficient Channel Attention (ECA) only, 4) Multi-Scale Spatial Attention (MSSA) only, 5) ECA combined with MSSA, 6) ECA with adaptive weights, 7) MSSA with adaptive weights, and 8) the full Enhanced CBAM framework. This progressive evaluation isolates the impact of each innovation, assessing their individual and synergistic effects on recognition performance. Table 2 details the performance metrics for each configuration.

The ECA mechanism yields a 0.2% accuracy gain on UT-HAR while reducing parameters by 4.3% compared with conventional MLP-based channel attention, highlighting its parameter efficiency. Multi-scale spatial

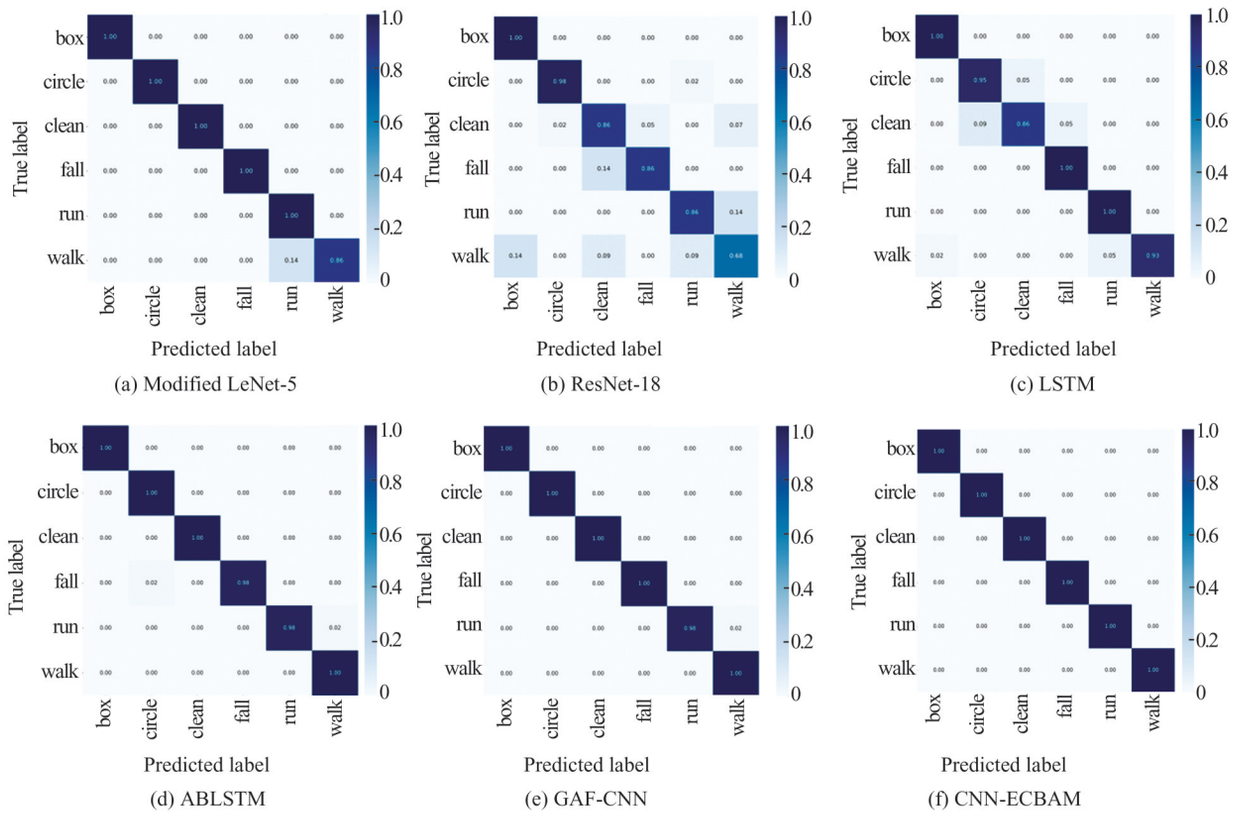


Fig.11 Confusion matrices for all evaluated models on the NTU-Fi_HAR dataset

Table 2 Enhanced CBAM component ablation results

Configuration	Acc./%		Params /10^6	FLOPs /10^9
	UT-HAR	NTU-Fi_HAR		
CNN without attention	95.8	97.5	2.1	0.8
Original CBAM（Baseline）	98.6	99.4	2.3	0.9
+ ECA Only	98.8	99.5	2.2	0.85
+ MSSA Only	98.9	99.6	2.4	1.0
+ ECA + MSSA	99.0	99.7	2.5	1.05
+ ECA + Adaptive Weights	99.0	99.7	2.3	0.9
+ MSSA+ Adaptive Weights	99.1	99.8	2.5	1.1
Full Enhanced CBAM (Ours)	**99.2**	**100**	**2.6**	**1.15**

Note：The best results are highlighted in bold.

attention adds 0.3%, underscoring the importance of capturing spatial patterns across multiple receptive field sizes, while learnable fusion weights further enhance performance by dynamically balancing channel and spatial attention. Overall, the enhanced CBAM achieves a 0.6% improvement over the standard CBAM, with synergistic rather than purely additive effects. All improvements are statistically significant ($p < 0.05$) across 10 independent runs.

5 Conclusion

A novel framework, CNN-ECBAM, is proposed for WiFi-based human activity recognition to address the inherent challenges of CSI-based sensing. First, a CSI-to-pseudo-color image conversion method is developed to preserve essential signal characteristics for effective CNN processing. Second, an enhanced CBAM architecture is introduced with improved channel and spatial attention mechanisms tailored to CSI data. Third, an integrated CNN framework is constructed in which these components are seamlessly combined. Extensive experiments on two public datasets demonstrate that CNN-ECBAM achieves competitive performance compared with existing methods, showing consistent accuracy improvements on both UT-HAR and NTU-Fi_HAR datasets. The enhanced attention mechanisms allow more effective capture of complex CSI patterns, resulting in higher recognition accuracy and robustness, thereby supporting deployment in real-world scenarios such as smart homes, healthcare monitoring, and building man-

agement.

Future work will focus on the design of lightweight architectures for edge deployment, enhancement of cross-domain generalization, and integration of multi-modal sensing to further improve recognition performance in diverse environments.

References

[1] Tao Z, Guo J, Liu Y. Efficient human behavior recognition method based on multi-antenna decision CSI[J]. *Journal of Computer Science and Exploration*, 2021, **15**: 1122-1132.

[2] Yang X L, He A L, Zhou M, *et al*. Human activity recognition system based on channel state information[C]//2018 *IEEE* 8*th Annual International Conference on CYBER Technology in Automation*, *Control*, *and Intelligent Systems* (*CYBER*). New York: IEEE, 2019: 1415-1420.

[3] Ma J Y, Wang H, Zhang D Q, *et al*. A survey on Wi-Fi based contactless activity recognition[C]//2016 *Intl IEEE Conferences on Ubiquitous Intelligence & Computing*, *Advanced and Trusted Computing*, *Scalable Computing and Communications, Cloud and Big Data Computing*, *Internet of People*, *and Smart World Congress* (*UIC*/*ATC*/*ScalCom*/*CBDCom*/*IoP*/*SmartWorld*). New York: IEEE, 2017: 1086-1091.

[4] Zou H, Yang J F, Das H P, *et al*. WiFi and vision multimodal learning for accurate and robust device-free human activity recognition[C]//2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*). New York: IEEE, 2020: 426-433.

[5] Miao F C, Huang Y X, Lu Z Y, *et al*. Wi-Fi sensing techniques for human activity recognition: Brief survey, potential challenges, and research directions[J]. *ACM Computing Surveys*, 2025, **57**(5): 1-30.

[6] Wang H, Zhang D Q, Wang Y S, *et al*. RT-fall: A real-time and contactless fall detection system with commodity WiFi devices[J]. *IEEE Transactions on Mobile Computing*, 2017, **16**(2): 511-526.

[7] Zou H, Zhou Y X, Yang J F, *et al*. WiFi-enabled device-free gesture recognition for smart home automation[C]//2018 *IEEE* 14*th International Conference on Control and Automation* (*ICCA*). New York: IEEE, 2018: 476-481.

[8] Duan P S, Diao X G, Cao Y J, *et al*. A comprehensive survey on Wi-Fi sensing for human identity recognition[J]. *Electronics*, 2023, **12**(23): 4858.

[9] Wang D Z, Yang J F, Cui W, *et al*. CAUTION: A robust WiFi-based human authentication system via few-shot open-set recognition[J]. *IEEE Internet of Things Journal*, 2022, **9** (18): 17323-17333.

[10] Zou H, Zhou Y X, Yang J F, *et al*. FreeCount: Device-free crowd counting with commodity WiFi[C]//*GLOBECOM* 2017 — 2017 *IEEE Global Communications Conference*. New York: IEEE, 2018: 1-6.

[11] Youssef M, Mah M, Agrawala A. Challenges: Device-free passive localization for wireless environments[C]//*Proceedings of the* 13*th Annual ACM International Conference on Mobile Computing and Networking*. New York: ACM, 2007: 222-229.

[12] Patwari N, Wilson J. RF sensor networks for device-free localization: Measurements, models, and algorithms[J]. *Proceedings of the IEEE*, 2010, **98**(11): 1961-1973.

[13] Wang Y, Liu J, Chen Y Y, *et al*. E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures[C]//*Proceedings of the* 20*th Annual International Conference on Mobile Computing and Networking*. New York: ACM, 2014: 617-628.

[14] Wang W, Liu A X, Shahzad M. Gait recognition using WiFi signals[C]//*Proceedings of the* 2016 *ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York: ACM, 2016: 363-373.

[15] Zeng Y Z, Pathak P H, Mohapatra P. WiWho: WiFi-based person identification in smart spaces[C]//2016 15*th ACM/IEEE International Conference on Information Processing in Sensor Networks* (*IPSN*). New York: IEEE, 2016: 1-12.

[16] Zhang D Q, Wang H, Wu D. Toward centimeter-scale human activity sensing with Wi-Fi signals[J]. *Computer*, 2017, **50** (1): 48-57.

[17] Yousefi S, Narui H, Dayal S, *et al*. A survey on behavior recognition using WiFi channel state information[J]. *IEEE Communications Magazine*, 2017, **55**(10): 98-104.

[18] Moshiri P F, Nabati M, Shahbazian R, *et al*. CSI-based human activity recognition using convolutional neural networks [C]//2021 11*th International Conference on Computer Engineering and Knowledge* (*ICCKE*). New York: IEEE, 2022: 7-12.

[19] Jiao W G, Zhang C S. An efficient human activity recognition system using WiFi channel state information[J]. *IEEE Systems Journal*, 2023, **17**(4): 6687-6690.

[20] Zhuravchak A, Kapshii O, Pournaras E. Human activity recognition based on Wi-Fi CSI data—A deep neural network approach[J]. *Procedia Computer Science*, 2022, **198**: 59-66.

[21] Guo L L, Zhang H, Wang C, *et al*. Towards CSI-based diversity activity recognition via LSTM-CNN encoder-decoder neural network[J]. *Neurocomputing*, 2021, **444**: 260-273.

[22] Zou H, Zhou Y X, Yang J F, *et al*. DeepSense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network[C]//2018 *IEEE International Conference on Communications* (*ICC*). New York: IEEE, 2018: 1-6.

[23] Ding X, Jiang T, Zhong Y, *et al*. Wi-Fi-based location-independent human activity recognition with attention mechanism enhanced method[J]. *Electronics*, 2022, **11** (4): 642.

[24] Mekruksavanich S, Phaphan W, Hnoohom N, *et al*. Attention-based hybrid deep learning network for human activity recognition using WiFi channel state information[J]. *Applied Sciences*, 2023, **13**(15): 8884.

[25] Luo F, Khan S, Jiang B, *et al*. Vision transformers for human activity recognition using WiFi channel state information[J]. *IEEE Internet of Things Journal*, 2024, **11**(17): 28111-28122.

[26] Li B, Cui W, Wang W, *et al*. Two-stream convolution augmented transformer for human activity recognition[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, **35**(1): 286-293.

[27] Hussain A, Chen Y S, Ullah A, *et al*. WiSigPro: Transformer for elevating CSI-based human activity recognition through attention mechanisms[J]. *Expert Systems with Applications*, 2024, **258**: 124976.

[28] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[EB/OL]. [2014-07-13]. https://arxiv.org/abs/1409.0473.

[29] Luong T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[C]//*Proceedings of the* 2015 *Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2015: 1412-1421.

[30] Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, **30**: 5998-6008.

[31] Wang F, Jiang M Q, Qian C, *et al*. Residual attention network for image classification[C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). New York: IEEE, 2017: 6450-6458.

[32] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2018: 7132-7141.

[33] Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. [2020-11-18]. https://arxiv. org/abs/ 2010.11929.

[34] Woo S, Park J, Lee J Y, *et al*. CBAM: Convolutional block attention module[M]//*Computer Vision – ECCV* 2018. Cham: Springer International Publishing, 2018: 3-19.

[35] Qin Y, Song D J, Cheng H F, *et al*. A dual-stage attention-based recurrent neural network for time series prediction[C]// *Proceedings of the* 26*th International Joint Conference on Artificial Intelligence*. New York: ACM, 2017: 2627-2633.

[36] Song H, Rajan D, Thiagarajan J J, *et al*. Attend and diagnose: Clinical time series analysis using attention models [C]// *Proceedings of the AAAI Conference on Artificial Intelligence.* New Orleans: AAAI Press, 2018: 4091-4098.

[37] Chorowski J K, Bahdanau D, Serdyuk D, *et al*. Attention-based models for speech recognition[C]//*Advances in Neural Information Processing Systems* 28 (*NIPS* 2015). Montreal: NIPS, 2015: 577-585.

[38] Bahdanau D, Chorowski J, Serdyuk D, *et al*. End-to-end attention-based large vocabulary speech recognition[C]// 2016 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*). New York: IEEE, 2016: 4945-4949.

[39] Kong Q Q, Cao Y, Iqbal T, *et al*. PANNs: Large-scale pretrained audio neural networks for audio pattern recognition [J]. *IEEE/ACM Transactions on Audio*, *Speech*, *and Language Processing*, 2020, **28**: 2880-2894.

[40] Li H, Yang W, Wang J X, *et al*. WiFinger: Talk to your smart devices with finger-grained gesture[C]//*Proceedings of the* 2016 *ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York: ACM, 2016: 250-261.

[41] Chen Z H, Zhang L, Jiang C Y, *et al*. WiFi CSI based passive human activity recognition using attention based BLSTM[J]. *IEEE Transactions on Mobile Computing*, 2019, **18**(11): 2714-2724.

[42] Shi W G, Tang Y F, Cao Y, *et al*. CIT-HAR: A high accuracy and lightweight human activity recognition system using CSI heatmaps and a hybrid transformer network[J]. *IEEE Transactions on Instrumentation and Measurement*, 2025, **74**: 2541317.

[43] Lu J, Batra D, Parikh D, *et al*. ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks[C]//*Advances in Neural Information Processing Systems* 32 (*NeurIPS* 2019). Vancouver: NeurIPS, 2019: 13-23.

[44] Nagrani A, Yang S, Arnab A, *et al*. Attention bottlenecks for multimodal fusion[C]//*Advances in Neural Information Processing Systems* 34 (*NeurIPS* 2021). Vancouver: NeurIPS, 2021: 14200-14213.

[45] LeCun Y, Boser B, Denker J S, *et al*. Backpropagation applied to handwritten ZIP code recognition[J]. *Neural Compu-*

tation, 1989, **1**(4): 541-551.

[46] LeCun Y, Bottou L, Bengio Y, *et al*. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, **86**(11): 2278-2324.

[47] Wang Q L, Wu B G, Zhu P F, *et al*. ECA-net: Efficient channel attention for deep convolutional neural networks[C]// 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). New York: IEEE, 2020: 11531-11539.

[48] Yang J F, Chen X Y, Zou H, *et al*. EfficientFi: Toward large-scale lightweight WiFi sensing *via* CSI compression[J]. *IEEE Internet of Things Journal*, 2022, **9**(15): 13086-13095.

[49] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition[C]//2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). New York: IEEE, 2016: 770-778.

基于增强型CBAM与卷积神经网络的人体活动识别方法

胡必玲[1,2]， 仝钰[1]

1.合肥师范学院 计算机与人工智能学院,安徽 合肥 236032

2.安徽省运动健康信息监测技术工程研究中心(合肥师范学院),安徽 合肥 236032

摘要：基于WiFi的人体活动识别为普适监测提供了一种非侵入式方法，然而同时实现高精度和高鲁棒性的监测仍然是一个重大挑战。本文提出一种基于增强型卷积块注意力机制的卷积神经网络框架（CNN-ECBAM），将原始信道状态信息系统性地转换为伪彩色图像，有效保留了深度神经网络处理所需的重要信号特征。核心创新在于设计了一种针对CSI数据特性的增强型卷积块注意力模块（ECBAM），该模块融合了高效通道注意力（ECA）与多尺度空间注意力（MSA），并通过可学习的自适应融合权重实现两者的动态协同，使网络能够捕获更具区分性的空间与通道特征。ECBAM模块被集成到统一的卷积神经网络（CNN）中，形成整体的CNN-ECBAM模型。在UT-HAR和NTU-Fi_HAR数据集上的实验结果表明，CNN-ECBAM在识别准确率上取得了具有竞争力的性能，并超越了主流基准模型。在UT-HAR上准确率达99.2%（高于ResNet-18的98.6%），在NTU-Fi_HAR上准确率达100%（高于GAF-CNN的99.62%）。这些结果验证了该方法在高精度、高可靠WiFi-HAR中的有效性。

关键词：人体活动识别；深度学习；信道状态信息；ECBAM；伪彩色图像

□