CBAM-Enhanced ResNet-18 for Noise-Resilient Radar-based Human Activity Recognition in Assisted Living

Tayyaba Parveen[1], Rehan Khan[1], Umer Saeed[2], Insoo Koo[1]

[1]Department of Electric/Electronic and Computer Engineering, University of Ulsan, South Korea

[2]School of Computer Science and Technology, University of Bedfordshire, Luton, UK

(mTayyaba@mail.ulsan.ac.kr, rehan77@mail.ulsan.ac.kr, Umer.Saeed@beds.ac.uk, iskoo@ulsan.ac.kr)

Abstract

Radar-based Human Activity Recognition (HAR) enables contactless sensing for healthcare monitoring, particularly in elderly care and assisted living, where privacy and continuous observation are crucial. Human motion is characterized through micro-Doppler signatures derived from reflected radio-frequency signals, eliminating the need for cameras or wearable devices. Although deep learning approaches achieve high accuracy, evaluations are typically limited to controlled laboratory environments. In real deployments, performance of models is affected and the signal quality can be degraded by environmental noise and propagation effects. This study examines the robustness of radar-based HAR using frequency-modulated continuous-wave radar micro-Doppler spectrograms for six daily activities (including falls) by comparing a baseline ResNet-18 with an attention-enhanced ResNet-18 incorporating convolutional block attention module.

I. Introduction

Human Activity Recognition (HAR) identifies various human motions from sensor data. It has several potential uses including healthcare monitoring, elderly care, safety systems and home automation. In medical settings, early detection of anomalous activities, especially falls, is critical for reducing injury severity and enabling rapid assistance. Continuous monitoring of daily routines is therefore essential for safe independent or assisted living. Traditional HAR methods generally rely on vision or wearable sensors. However, these have disadvantages including limited battery life, privacy concerns and sensitivity to illumination levels. These constraints have increased interest in contactless sensing solutions. Radar-based sensing offers a privacy-preserving alternative by analyzing radio-frequency reflections and micro-Doppler signatures. This technology works under poor lighting and partial obstruction. The importance of such systems is underscored by World Health Organization projections indicating that by 2050, over one-third of the global population will be aged 65 years or older [1]. Although recent studies apply deep learning (DL) to radar-based activity recognition, most assume ideal signal conditions. This work investigates six daily life activities using frequency-modulated continuous-wave (FMCW) radar data to evaluate noise robustness and attention-based recognition reliability [2].

II. Proposed Methodology

This section describes the methodology adopted in this study, as illustrated in Fig. 1, including the radar dataset, preprocessing steps, baseline, attention-enhanced architectures and the evaluation protocol used to assess robustness under realistic noise conditions.

Radar Dataset and Preprocessing

Experiments are conducted using a publicly accessible FMCW radar-based HAR dataset consisting of micro-Doppler spectrogram images. The dataset contains recordings from 99 participants aged 21 to 99 years, providing significant subject diversity and motion variability. Six activities of daily living are analyzed: bending, drinking, falling, sitting, standing and walking. Each activity is recorded for a fixed duration; walking sequences are recorded longer to capture continuous motion [1].

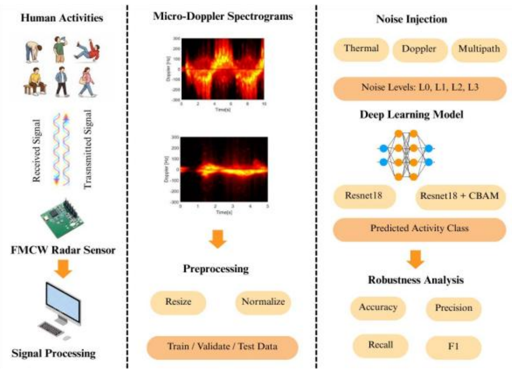

Fig. 1: System architecture for noise-robust radar-based human activity recognition.

Raw radar signals are processed using short-time Fourier transform techniques to generate time-frequency micro-Doppler spectrograms. These representations encode motion-induced Doppler frequency shifts from different body parts, producing distinctive patterns across activities. To ensure compatibility with convolutional neural networks pre-trained on natural images, all spectrograms are resized to 224 × 224 pixels and normalized using ImageNet mean and standard deviation values. For reproducibility and fair evaluation, the dataset is divided into training (50%), validation (25%) and testing (25%) subsets using stratified splitting to preserve class balance. The test set is fixed and used for all experiments.

Baseline and Attention-Enhanced Models

ResNet-18 is used as the baseline architecture for efficient residual learning and strong generalization. The network is initialized with ImageNet pre-trained weights and the final fully connected layer is adapted to classify six activities. Training uses only clean spectrograms with cross-entropy loss and the Adam optimizer at a learning rate of 1×10^{-4} for 20 epochs. Model selection relies on validation accuracy, retaining the best checkpoint. To improve noise robustness, Convolutional Block Attention Modules (CBAM) are embedded into the ResNet-18 backbone, applying channel and spatial attention after residual stages with 64, 128, 256 and 512 feature channels. The attention-enhanced model is trained under identical conditions as the baseline to ensure fair comparison.

Noise Modeling and Robustness Evaluation

Although the dataset is collected under controlled conditions, real radar signals are affected by a variety of noise sources. Three types of synthetic noise are applied at test time [3]. Thermal noise is modeled as additive white Gaussian noise:

$$X_{th}(i,j)=X(i,j)+N(0,\sigma^2) \tag{1}$$

where $X(i,j)$ represents a clean spectrogram and σ controls noise intensity. Phase noise and Doppler smearing are simulated using two-dimensional Gaussian blurring:

$$X_{blur}=X * G(\sigma_b,k) \tag{2}$$

Where $G(\sigma_b,k)$ is a Gaussian kernel with standard deviation σ_b and kernel size k. Multipath effects are approximated by adding a shifted and attenuated copy of the spectrogram:

$$X_{mp}(i,j)=X(i,j)+\alpha X(i\text{-}\Delta_i,j\text{-}\Delta_j) \tag{3}$$

where α controls reflection strength and Δi, Δj denote time and frequency shifts. The final noisy spectrogram is generated by sequentially combining all noise sources:

$$X_{noisy}=((X * G)+\alpha X_{shift})+N(0,\sigma^2) \tag{4}$$

Four noise levels are defined: L0 (clean), L1 (low), L2 (medium) and L3 (strong). Accuracy, precision, recall and macro-averaged F1-score are used as performance metrics to assess both models on the same fixed test set at all noise levels [4, 5].

III. Results and Discussions

The robustness of both models under increasing noise is demonstrated in Fig. 2, presenting accuracy trends from clean (L0) to high-noise (L3) settings. At L0, ResNet-18 achieves 95.68% accuracy, whereas the attention-enhanced model reaches 98.15%, indicating better feature representation. With rising noise, performance degrades for both models; however, the attention-based model consistently outperforms the baseline. At L3, accuracy drops to 70.99% for ResNet-18 but stays at 75.93% with CBAM-enhanced, indicating improved resilience. Table 1 further confirms widening performance gaps across metrics.

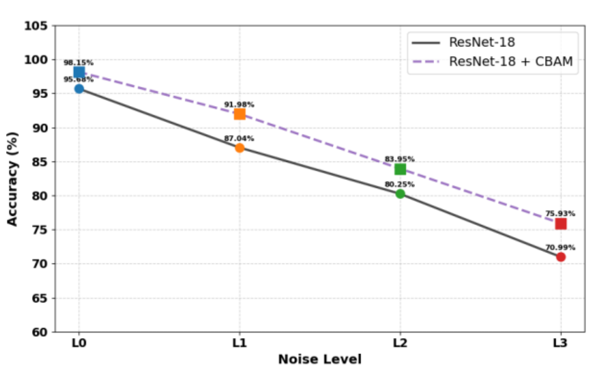

Fig. 2: Classification accuracy of ResNet-18 and ResNet-18 with CBAM under different noise levels.

Table 1: Robustness comparison of baseline and CBAM enhanced ResNet-18.

Noise Level	Model	Acc. (%)	Prec. (%)	Recall (%)	F1-Score (%)
L0 (Clean)	ResNet-18	95.68	96.06	95.68	95.70
	ResNet-18 + CBAM	98.15	98.21	98.15	98.15
L1 (Low)	ResNet-18	87.04	92.71	87.04	87.73
	ResNet-18 + CBAM	91.98	94.28	91.98	92.19
L2 (Medium)	ResNet-18	80.25	88.19	80.25	80.62
	ResNet-18 + CBAM	83.95	90.67	83.95	84.58
L3 (High)	ResNet-18	70.99	83.73	70.99	71.75
	ResNet-18 + CBAM	75.93	86.30	75.93	76.00

IV. Conclusion

This study evaluated the robustness of DL-based radar HAR under realistic noise conditions relevant to healthcare monitoring. Using FMCW radar micro-Doppler spectrograms for six daily activities, both a baseline ResNet-18 and an attention-enhanced model were tested across multiple noise levels. Performance declined as noise severity increased, highlighting the limitations of evaluations based only on clean data. However, the attention-enhanced model consistently outperformed the baseline. Accuracy improved from 95.68% to 98.15% under clean conditions and from 70.99% to 75.93% under high noise, with corresponding gains in F1-score. These results show that attention mechanisms improve robustness to signal degradation and enhance recognition reliability, supporting their suitability for real-world, noise-prone healthcare monitoring scenarios.

ACKNOWLEDGMENT

This result was supported by the "Regional Innovation System & Education (RISE)" through the Ulsan RISE Center, funded by the Ministry of Education (MOE) and the Ulsan Metropolitan City, Republic of Korea.(2025-RISE-07-001)

REFERENCES

1. Saeed, U., et al., Discrete human activity recognition and fall detection by combining FMCW RADAR data of heterogeneous environments for independent assistive living. Electronics, 2021. 10(18): p. 2237.
2. Ding, C., et al., Human motion recognition with spatial-temporal-convLSTM network using dynamic range-doppler frames based on portable FMCW radar. IEEE Transactions on Microwave Theory and Techniques, 2022. 70(11): p. 5029-5038.
3. Khan, R., U. Saeed, and I. Koo, Robust Sensor Fault Detection in Wireless Sensor Networks Using a Hybrid Conditional Generative Adversarial Networks and Convolutional Autoencoder. IEEE Sensors Journal, 2025.
4. Nguyen, N., et al., Improving human activity classification based on micro-doppler signatures of FMCW radar with the effect of noise. PloS one, 2024. 19(8): p. e0308045.
5. Khan, R., U. Saeed, and I. Koo, FedLSTM: A federated learning framework for sensor fault detection in wireless sensor networks. Electronics, 2024. 13(24): p. 4907.