Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD29440

SNAP-HAR: Signal-Neural Adaptive Processing for Robust Human Activity Recognition

Huazhen Liu

Branksome Hall, Toronto, Canada
jennyliu080306@gmail.com

Abstract. Human Activity Recognition (HAR) using wearable sensors is essential for healthcare and smart home applications, yet real-world deployment remains challenging due to sensor noise. Current self-supervised methods apply uniform architectures regardless of signal quality, leading to poor performance on noisy data. This paper presents SNAP-HAR, a framework implementing Signal-Neural Adaptive Processing that jointly optimizes signal preprocessing and neural architectures based on dataset-specific noise characteristics. Power spectral density analyses quantify orders-of-magnitude differences between laboratory-preprocessed (UCI-HAR) and real-world (USC-HAD, MotionSense) datasets. Our adaptive processing achieves consistent improvements across all configurations, with gains up to 17.0% on clean data and 24.6% on noisy conditions. Most significantly, SNAP-HAR elevates real-world performance to 0.9121 F1-score, matching the 0.9276 laboratory baseline. This convergence validates that robust HAR is achievable through adaptive signal-neural processing, eliminating the deployment gap that historically limited practical applications. Universal improvements confirm our approach provides architecture-agnostic enhancements applicable to self-supervised paradigms.

Keywords: Human Activity Recognition, Signal Denoising, Self-supervised Learning, Dataset Adaptation, Noise Robustness

1. Introduction

Human activity recognition (HAR) is a computational task that infers daily activities from sensor data. HAR models transform noisy time series into semantic activity labels in real time. This field has emerged as critical for numerous applications including healthcare monitoring, workplace safety and ergonomics, and smart home automation and interfaces [1]. Among available sensing modalities, inertial measurement units (IMUs) have become the standard deployment choice for HAR due to their ubiquity in smartphones and wearables.

HAR methodology has evolved through distinct phases, beginning with classical machine learning techniques and advancing to deep learning. Early approaches employed SVMs for smartphone data [2] and random forests for feature extraction. The field was transformed by deep learning, starting with CNNs that learn features directly from raw sensor data [3] and advancing to hybrid architectures like DeepConvLSTM for capturing local patterns and long-range dependencies [4]. Recent advances have shifted to transformers and self-supervised learning. Pioneering masked reconstruction with ConvTransformers established self-supervised pretraining [5], and Wang et al. advanced this by showing channel-aware masking provides the strongest representations for label-efficient HAR [6].

Despite these advances, noise remains a critical obstacle for real-world deployment. Sensor imperfections generate various noise types including drift over time, quantization errors, and motion artifacts from inconsistent device placement [7]. These disturbances distort accelerometer and gyroscope readings, interfering with activity discrimination particularly for subtle or overlapping movements. Performance varies substantially across datasets, with baseline methods achieving F1-scores of 0.9276 on preprocessed UCI-HAR but dropping to 0.6220 on USC-HAD and 0.8493 on MotionSense even in best configurations [6]. This performance degradation reflects fundamental differences in data collection, as UCI-HAR employs controlled laboratory conditions with standardized protocols and Butterworth preprocessing, while USC-HAD and MotionSense capture naturalistic settings with variable sensor placement and unprocessed signals [8].

To address these challenges, we present SNAP-HAR, a comprehensive framework that combines signal processing insights with neural architecture optimization. Our approach introduces adaptive signal processing with noise-calibrated augmentation including Mixup, time warping, and magnitude scaling, where intensity scales proportionally to measured spectral noise levels. We systematically modify neural architectures through normalization scheme adaptation, activation function selection, and regularization tuning optimized for specific noise profiles. Additionally, we achieve 62.7% parameter reduction through unstructured L1 pruning while maintaining superior performance. These contributions collectively enable MotionSense to achieve 0.9121 F1-score, matching the baseline performance of preprocessed UCI-HAR data and demonstrating that principled adaptation bridges the laboratory to deployment gap.

2. Related work

2.1. Self-supervised learning for HAR

Self-supervised learning (SSL) learns representations from unlabeled data, proving essential when annotations are costly. In HAR, masked reconstruction pretraining followed by fine-tuning has emerged as the leading strategy [5]. Models pretrain by reconstructing randomly masked sensor sequences, then fine-tune on limited labeled data, significantly reducing annotation requirements while maintaining competitive performance. Wang et al. advanced this paradigm through channel-aware masking, demonstrating that considering sensor channel structure during masking provides stronger representations than temporal masking alone [6]. Their work established the current state-of-the-art for self-supervised HAR, motivating our investigation into noise-adaptive enhancements.

2.2. Signal processing and attention mechanisms

Classical denoising remains fundamental to HAR preprocessing. Butterworth filters remove high-frequency noise beyond human movement ranges while preserving activity signatures [9]. Median filters eliminate impulse noise without distorting signal edges, crucial for transition detection. Hybrid approaches integrate these techniques with neural networks, combining signal conditioning with deep learning for end-to-end optimization. Attention mechanisms revolutionized HAR architectures through adaptive weighting across temporal and spatial dimensions. Channel attention via SENet [10] and CBAM [11] enables learning cross-modal sensor importance, emphasizing informative accelerometer/gyroscope channels while suppressing noisy ones.

2.3. Benchmark datasets

Three datasets exemplify the spectrum of data quality in HAR research. UCI-HAR [12] represents controlled data collection from 30 participants in laboratory settings, preprocessed with 20Hz Butterworth filtering, gravity-body separation, and fixed 2.56-second windows with 50% overlap. MotionSense [13] captures real-world smartphone usage from 24 participants without preprocessing, preserving noise from variable device placement and environmental factors. USC-HAD [14] bridges these extremes with 100Hz sampling from 14 participants using hip-mounted sensors, providing raw readings without signal decomposition and including 12 activity classes.

3. SNAP-HAR framework

3.1. Overview and architecture

We propose SNAP-HAR, a Signal-Neural Adaptive Processing framework that extends masked reconstruction-based self-supervised learning with noise-aware signal processing and architectural optimization. While our approach builds upon the channel-aware masking strategies introduced by Wang et al. [6], we fundamentally enhance the paradigm through adaptive preprocessing and architecture modifications tailored to dataset-specific noise characteristics.

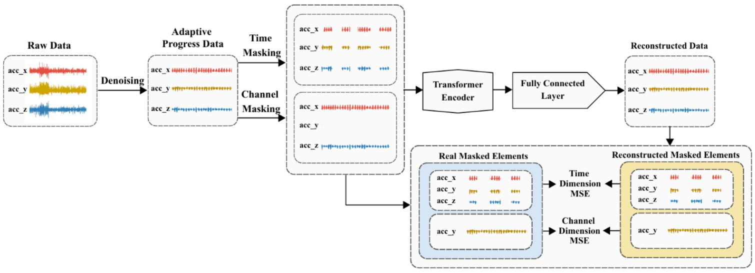

Figure 1. SNAP-HAR pretraining pipeline with MotionSense data. Denoised accelerometer signals undergo masking (time or channel masking shown as examples) before transformer-based reconstruction via MSE loss optimization

Figure 1 presents the SNAP-HAR pretraining pipeline using MotionSense data as an example. Raw sensor data first passes through a denoising stage with augmentation intensities adaptively calibrated based on spectral noise analysis, ensuring signal conditioning that aids reconstruction without compromising discriminative features. The adaptive calibration methodology detailed in Section 3.3 determines these dataset-specific parameters. Following denoising, we implement masked reconstruction where portions of the sensor sequence are hidden, forcing the model to learn robust representations by predicting missing elements. We evaluate five masking strategies (time, span, channel, time-channel, and span-channel), with Figure 1 demonstrating time and channel masking as representative examples. Time masking removes the same randomly selected timesteps across all sensor channels at a specified ratio while channel masking eliminates entire sensor axes such as one or two complete accelerometer channels to encourage cross-modal learning. The masked sequences pass through a transformer encoder with three layers, 128-dimensional embeddings, and

4-head attention mechanisms that reconstructs original values by minimizing MSE loss. After self-supervised pretraining captures underlying activity patterns, the frozen encoder connects to a classification head for supervised fine-tuning on labeled data.

3.2. Spectral analysis for noise characterization

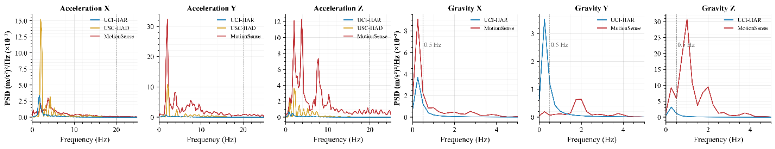

Figure 2. Power spectral density (PSD) analysis across triaxial accelerometer and gravity components for three datasets. Reference lines at 20 Hz and 0.5 Hz indicate physiological movement boundary and gravitational signal bandwidth respectively

The foundation of SNAP-HAR lies in understanding that different datasets exhibit fundamentally different noise characteristics, necessitating adaptive rather than uniform processing. Power spectral density (PSD) analysis quantifies signal quality through frequency-domain decomposition, revealing critical differences that directly calibrate our denoising parameters.

Figure 2 reveals noise disparities across datasets and sensor axes. For acceleration components, UCI-HAR maintains noise floors below $2.5\ (m/s^2)^2/Hz$ across all axes due to $20Hz$ Butterworth filtering. MotionSense shows 10-15x higher spectral power, with X-axis peaks reaching $15\ (m/s^2)^2/Hz$ at $8-10\ Hz$ harmonics and Y/Z axes exhibiting broadband elevation up to $30\ (m/s^2)^2/Hz$. USC-HAD demonstrates intermediate characteristics with $5\ (m/s^2)^2/Hz$ baseline but prominent spikes at $12\ Hz$ and $18\ Hz$ suggesting mechanical interference. The gravity components provide starker contrasts. UCI-HAR concentrates 95% of gravitational energy below $0.5\ Hz$ with peak magnitudes of $0.004\ (m/s^2)^2/Hz$, correctly isolating postural changes. MotionSense gravity estimates contain spurious high-frequency content extending to $3\ Hz$ with magnitudes reaching $0.03\ (m/s^2)^2/Hz$, representing 7.5x noise amplification corrupting posture discrimination. USC-HAD cannot perform gravity separation, limiting its ability to distinguish static postures from dynamic movements. These measured spectral characteristics directly determine our adaptive parameters, with noise power ratios translating to augmentation intensities.

3.3. Dataset-adaptive preprocessing

Our adaptive preprocessing calibrates augmentation intensities based on spectral characteristics quantified in Section 3.2. For time warping, we apply cubic spline interpolation with Gaussian perturbations $N(0,\ \sigma^2)$, where σ scales proportionally with measured noise levels. Clean datasets use $\sigma = 0.1$ to preserve signal integrity, while noisy datasets require $\sigma = 0.3$ for stronger regularization. This ensures augmentation strength matches inherent data quality.

Magnitude warping employs multiplicative scaling through Gaussian-filtered random fields, providing smooth amplitude variations while preserving activity patterns. Mixup regularization combines samples using coefficients λ sampled from $Beta(\alpha,\ \alpha)$ .The parameter α inversely correlates with noise level: clean data uses $\alpha = 2.0$ enabling aggressive mixing for regularization, while noisy signals use $\alpha = 0.5$ to prevent excessive corruption. These calibrated adaptations ensure augmentation enhances rather than degrades signal quality across diverse noise conditions.

3.4. Architecture optimization

SNAP-HAR introduces architectural modifications optimized for dataset-specific characteristics. For high-variance datasets like USC-HAD and MotionSense, we replace batch normalization with layer normalization, computing statistics within individual samples to achieve distributional invariance crucial for non-stationary noise. We substitute ReLU with GELU activation, defined as $GELU(x) = x \cdot \Phi(x)$ where Φ represents the standard normal CDF, enabling smooth gradient propagation through noisy regions.

Training enhancements include Stochastic Weight Averaging during the final 25% of epochs, guiding models toward flatter minima that generalize better to noisy conditions. Test-time augmentation averages predictions across ten augmented versions of each test sample. Through unstructured L1 pruning, we achieve 62.7% parameter reduction while maintaining superior performance, enabling deployment on resource-constrained edge devices.

4. Experiments

4.1. Experimental setup

We evaluate on three benchmarks representing noise spectra. Dataset splits follow established protocols: UCI-HAR (70/30 subject-wise), USC-HAD (subjects 1-10 training, 11-14 testing), MotionSense (80/20 split). Subject-wise divisions ensure generalization to unseen individuals. Macro-F1 serves as primary metric, crucial for USC-HAD's imbalanced 12-class structure.

4.2. Main results

Tables 1-3 demonstrate that SNAP-HAR achieves substantial improvements across all datasets and masking strategies, with gains closely tied to measured noise characteristics.

Table 1. UCI-HAR results (clean dataset)

Masking strategies	Baseline	CascadeHAR (Ours)	Improvement
Time Masking	0.7924 [5]	0.9274	+17.0%
Span Masking	0.8923 [15]	0.9309	+4.33%
Channel Masking	0.9226 [6]	0.9387	+1.75%
Time-channel Masking	0.9234 [6]	0.9324	+0.975%
Span-Channel Masking	0.9276 [6]	0.9345	+0.744%

Table 2. USC-HAD results (moderate noise dataset)

Masking strategies	Baseline	CascadeHAR (Ours)	Improvement
Time Masking	0.5039 [5]	0.5974	+18.6%
Span Masking	0.4932 [15]	0.6146	+24.6%
Channel Masking	0.6149 [6]	0.6167	+0.293%
Time-channel Masking	0.6220 [6]	0.6672	+7.27%
Span-Channel Masking	0.5825 [6]	0.6452	+10.8%

Table 3. MotionSense results (high noise dataset)

Masking strategies	Baseline	CascadeHAR (Ours)	Improvement
Time Masking	0.8279 [5]	0.9121	+8.42%
Span Masking	0.8317 [15]	0.8923	+6.06%
Channel Masking	0.8493 [6]	0.8640	+1.47%
Time-channel Masking	0.8428 [6]	0.8920	+4.92%
Span-Channel Masking	0.8396 [6]	0.9062	+6.66%

Our results reveal improvement magnitudes correlating directly with dataset noise characteristics, validating the adaptive processing approach. Wang et al.'s channel-aware masking already established superiority over traditional supervised methods across all datasets [6], our adaptive enhancements achieve substantial gains beyond their self-supervised baseline, establishing SNAP-HAR as the new state-of-the-art for noise-robust HAR. USC-HAD, with moderate noise and no gravity separation, achieves the highest overall improvement despite having 12 activity classes. Most remarkably, span masking improves from 0.4932 to 0.6146 F1-score, a 24.6% gain transforming a failing system into a viable solution. This dramatic enhancement occurs because USC-HAD's intermediate noise profile benefits maximally from calibrated augmentation providing sufficient denoising to stabilize learning without over-smoothing discriminative features. The lack of gravity separation, initially a fundamental limitation, becomes manageable through adaptive processing compensating for missing signal components.

MotionSense demonstrates our ability to bridge the deployment gap, with time masking achieving 0.9121 F1-score, matching UCI-HAR's baseline of 0.9276. This convergence is significant because MotionSense represents real-world smartphone data without preprocessing, while UCI-HAR underwent Butterworth filtering, gravity-body separation, and controlled collection. Consistent improvements across all masking strategies elevate noisy sensor data to laboratory-quality performance, proving adaptive processing can replace complex preprocessing pipelines. UCI-HAR shows meaningful improvements with time masking gaining 17.0%, suggesting adaptive augmentation provides regularization benefits beyond noise handling.

The mechanism behind these improvements becomes clear when examining masking strategy differences. Time and span masking are inherently vulnerable to noise, as temporal corruption directly interferes with reconstruction objectives. In noisy data, models cannot distinguish between intentional masks requiring prediction and inherent noise requiring filtering, causing baseline performance to collapse below 0.50 on USC-HAD. Our adaptive preprocessing addresses this by applying calibrated time warping that smooths noise while preserving activity dynamics, enabling focus on genuine reconstruction targets. Channel masking shows modest but consistent gains because sensor redundancy naturally mitigates single-channel noise. Architectural optimizations including layer normalization and GELU activation still provide measurable improvements by handling residual cross-channel correlations. These results establish that principled signal-neural co-design achieves what neither approach alone could accomplish.

4.3. Implementation details

PyTorch 2.0 implementation follows a two-phase paradigm: pretraining reconstructs masked sequences for 150 epochs using MSE loss, then fine-tuning trains only the classification head for 100-150 epochs using cross-entropy loss, keeping the encoder frozen. For UCI-HAR, USC-HAD, and MotionSense, batch sizes are 512, 256, and 256 respectively. Masking ratios use 0.15 for time/span and 0.5 for channel masking. Combined strategy loss weights α are 0.8, 0.7, and 0.8 respectively. Learning rates are 5e-4, 5e-4, and 8e-4 respectively. SNAP-HAR employs AdamW with 0.01 weight decay versus Adam. Noise-adaptive parameters scale with PSD levels: time warping σ ranges from 0.1 (clean) to 0.3 (noisy), while Mixup α inversely scales from 2.0 to 0.5. Stochastic Weight Averaging activates after 75% of epochs, test-time augmentation averages 10 predictions, and early stopping uses 20-epoch patience on validation F1. The framework achieves 62.7% parameter reduction through L1 pruning, enabling edge deployment.

5. Conclusion

SNAP-HAR addresses noise-induced performance degradation through signal-neural adaptive processing. Building upon masked reconstruction-based self-supervised learning, our framework introduces dataset-specific adaptations bridging laboratory and real-world deployment. Spectral analyses quantified orders-of-magnitude differences in noise characteristics, revealing the inadequacy of uniform approaches. The framework achieves consistent improvements across all configurations, with gains reaching 24.6% on noisy datasets. Significantly, SNAP-HAR elevates real-world MotionSense to 0.9121 F1-score, matching laboratory baseline (0.9276). This convergence proves robust HAR is achievable through adaptive processing, transforming HAR from laboratory-constrained research to deployable technology. Universal improvements across masking strategies confirm architecture-agnostic enhancements applicable to self-supervised paradigms. The 62.7% parameter reduction through pruning enables edge deployment.

Future work should explore automated noise profiling for zero-shot adaptation and investigate transferability to other sensor-based tasks. As wearable sensing becomes ubiquitous, methods guaranteeing robust performance across diverse conditions will be essential for HAR deployment.

References

[1] Zhou, B., Yang, J. and Li, Q. (2019) Smartphone-Based Activity Recognition for Indoor Localization Using a Convolutional Neural Network. Sensors, 19, 621.

[2] Anguita, D., Ghio, A., Oneto, L., Parra, X. and Reyes-Ortiz, J.L. (2012) Human Activity Recognition Using Support Vector Machines on Smartphones. Proceedings of IWAAL, 216-223.

[3] Ronao, C.A. and Cho, S.B. (2015) Deep Convolutional Neural Networks for Human Activity Recognition with Smartphone Sensors. Lecture Notes in Computer Science, 9492, 46-53.

[4] Ordóñez, F. and Roggen, D. (2016) Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors, 16, 115.

[5] Haresamudram, H., Beedu, A., Agrawal, V., Grady, P.L., Essa, I., Hoffman, J. and Plötz, T. (2020) Masked Reconstruction Based Self-Supervision for Human Activity Recognition. Proceedings of ISWC, 45-49.

[6] Wang, J., Zhu, T. and Ning, H. (2023) An Improved Masking Strategy for Self-supervised Masked Reconstruction in Human Activity Recognition. ArXiv Preprint.

[7] Kaur, H., Rani, V. and Kumar, M. (2024) Human Activity Recognition: A Comprehensive Review. Expert Systems, 41, e13680.

[8] Bulling, A., Blanke, U. and Schiele, B. (2014) A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors. ACM Computing Surveys, 46, 1-33.

[9] Maitre, J., Bouchard, K. and Gaboury, S. (2023) Data Filtering and Deep Learning for Enhanced Human Activity Recognition from UWB Radars. Journal of Ambient Intelligence and Humanized Computing, 14, 7845-7856.

[10] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. Proceedings of CVPR, 7132-7141.

[11] Woo, S., Park, J., Lee, J.Y. and Kweon, I. (2018) CBAM: Convolutional Block Attention Module. Proceedings of ECCV, 3-19.

[12] Anguita, D., Ghio, A., Oneto, L., Parra, X. and Reyes-Ortiz, J.L. (2013) A Public Domain Dataset for Human Activity Recognition Using Smartphones. Proceedings of ESANN, 437-442.

[13] Malekzadeh, M., Clegg, R.G., Cavallaro, A. and Haddadi, H. (2018) Protecting Sensory Data Against Sensitive Inferences. Proceedings of W-P2DS, 1-6.

[14] Zhang, M. and Sawchuk, A.A. (2012) USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition. Proceedings of PervasiveHealth, 1036-1043.

[15] Xu, H., Zhou, J., Tan, R., Li, M. and Shen, G. (2021) LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications. Proceedings of SenSys, 220-233.