# mV-IMU: mmWave-enabled Virtual Inertia Measurement Unit for High-fidelity Activities of Daily Living Monitoring

Zhi Zhang*, Chongxin Zhong*, Yuliang Fu*, Ying Chen†, Jinyuan Jia†, Huining Li*, Chenhan Xu*
*Department of Computer Science, North Carolina State University, USA
Email: {zzhan224, czhong4, yfu34, hli83, cxu34}@ncsu.edu
†College of Information Science and Technology, Pennsylvania State University, USA
Email: {yfc5578, jinyuan}@psu.edu

***Abstract*—Monitoring human motion through inertial metrics is vital for healthcare, rehabilitation, and activity recognition. Traditional approaches rely on wearable inertial measurement units (IMUs), which, despite their accuracy, impose burdens due to their intrusive nature, limiting long-term usability. To mitigate this, recent advances explore device-free alternatives, such as pose-based inertial inference from video or mmWave sensing. However, inertial signals derived from pose tracking are prone to error amplification during differentiation. In this paper, we present mV-IMU, a novel mmWave-enabled Virtual Inertial Measurement Unit framework that bypasses pose estimation altogether to directly reconstruct body accelerations from raw mmWave signals. Our approach leverages a deep inertia reconstruction model trained on kinematics-informed features extracted from mmWave point clouds, integrated with a physics-guided optimization scheme for enhanced accuracy. Extensive evaluations show that mV-IMU achieves inertial measurement fidelity close to wearable IMUs, enabling practical, non-intrusive motion monitoring for smart healthcare and rehabilitation contexts.**

***Index Terms*—inertial measurement unit, mmWave sensing**

## I. Introduction

Monitoring activities of daily living (ADL) is crucial for smart healthcare, remote rehabilitation, and assistive diagnostics. Among various motion metrics, inertial measurement of body segments reflects joint loading, muscular effort, and energy expenditure, providing actionable insights for clinical decision-making [1]. Traditionally, such measurements rely on wearable inertial measurement units (IMUs). While IMUs offer high-fidelity data, their reliance on multiple body-worn sensors, frequent charging, and maintenance makes them intrusive to daily life, reducing comfort and adherence, and hindering long-term or continuous evaluation, particularly for those with mobility impairments.

To address these issues, researchers have explored device-free inertia measurement methods. Video-based frameworks, such as IMUTube [2], estimate inertial signals utilizing poses extracted from RGB videos of activities of daily living. However, cameras raise privacy concerns and are sensitive to lighting conditions. Sharing the same pipeline, other device-free pose tracking technologies (e.g., WiFi [3]) are also feasible of supporting inertial measurements. Among these methods, millimeter wave sensors recently achieved considerably high precision on human pose tracking [4] leveraging its short wavelength and high bandwidth. Nevertheless, we found inertial measurement based the precise poses tracked by mmWave is still of large errors (Sec. II), which is insufficient to bridge the existing wisdom established on wearable IMUs.

In this paper, we rethink and identify the fundamental issue with the mmWave-based inertia measurement based on pose-tracking (Sec. II). Specifically, inertial measurements are derived from pose. Even minor errors and noises in estimated poses will be magnified through differencing, resulting in unstable and unreliable inertia estimation. Therefore, we present mmWave-enabled Virtual Inertia Measurement Units (mV-IMU) that reconstructs human body accelerations directly from mmWave signals (Sec. III). Our method skips pose tracking and numerical differentiation to avoid error accumulation. Our detailed contributions are listed as follows:

- We present a new mmWave-based non-contact inertial measurement paradigm. The new paradigm can precisely measure inertia directly from mmWave signals, which avoid the large error introduced by pose tracking.
- We developed mV-IMU, a virtual inertial measurement unit system that does not require on-body sensors. mV-IMU is built upon inertia features extracted from mmWave point clouds, a high-fidelity deep inertia reconstruction model, and a Kinamatics-guide optimization to measure precise inertia.
- We comprehensively evaluated mV-IMU and demonstrate that mmWave-based virtual IMUs can achieve measurement fidelity close to wearable sensors, enabling practical, device-free acceleration monitoring for healthcare applications.

## II. Rethinking the mmWave Inertia Measurement

**Inertia estimation from RF human pose:** Recent developments in millimeter wave (mmWave) sensing have enabled accurate tracking of human pose and recognition of activity using radio frequency (RF) signals. Most related works focus on reconstructing joint positions or classifying actions, with little attention given to the direct inertia measurement such as acceleration. Intuitively, acceleration can be derived by applying numerical differentiation, such as the central difference method, to the sequence of estimated joint positions: $v_i = (x_{i+1} - x_{i-1})/(2\Delta t)$, $a_i = (x_{i+1} - 2x_i + x_{i-1})/\Delta t^2$, where $x_i$ is the joint position in the frame $i$ and $\Delta t$ is the sampling interval. However, *inertia estimated from RF human poses have large errors*, as we demonstrated below.

**Estimation error and analysis:** We evaluated this limitation using synchronized mmWave radar and camera-based pose data from the public `mRI` dataset [5]. Although both radar and camera provide generally consistent joint positions, accelerations computed by numerical differentiation diverge significantly between the two modalities. As shown in Fig. 1, the resulting acceleration curves not only differ in amplitude from the camera reference but also show spurious peaks and baseline drift. The lower frame rate of the radar further widens the sampling interval $\Delta t$, increasing the finite difference error and worsening the accuracy. These results show that pose differentiation alone cannot deliver reliable acceleration estimates. The reason is that acceleration is the second derivative of position. Even minor errors and noises in estimated joint positions are magnified through two rounds of differencing, resulting in unstable and unreliable acceleration estimates.

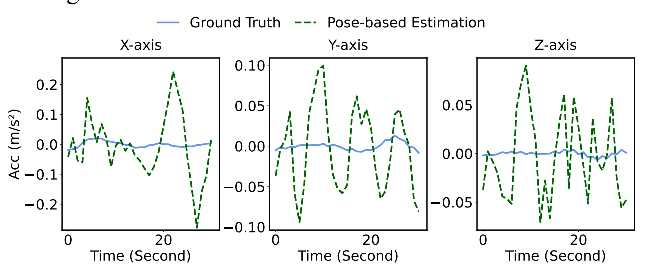


Fig. 1. Comparison of knee joint acceleration estimated from mmWave-derived pose and ground truth.

**Summary:** These findings highlight the need for a more reliable solution estimating the acceleration of human pose from mmWave sensing data. To address this challenge, we propose a Virtual-IMU framework that learns to directly estimate body segment accelerations from raw radar point clouds. By bypassing the error-prone differentiation step, our approach improves robustness and accuracy, enabling unobtrusive, device-free monitoring of biomechanical accelerations for clinical and daily life applications.

## III. METHODOLOGY

### A. Inertial Feature Extraction from mmWave Signals

To enable accurate acceleration estimation from mmWave signals, it is essential to select a data representation that preserves key motion information in ADL. Most current mmWave-based pose estimation methods use Range-Doppler-Angle (RDA) heatmaps generated through FFT-based spectral processing [6]. Although these heatmaps contain detailed information on range, speed, and angle of arrival, they also include a large amount of redundant and irrelevant data such as environmental clutter and noises. These interferences and redundancy deteriorate signal-to-noise ratio (SNR), which increases the complexity of models and reduces inference efficiency.

Therefore, our approach leverages mmWave point clouds [7] as features, which provide a direct and physically meaningful representation of motion. Point clouds are obtained from RDA through Constant False Alarm Rate (CFAR) denoising and spatial filtering. This process effectively removes environmental noise and clutter, ensuring that the extracted features remain informative for the following inertia reconstruction models. Formally, each mmWave sampling produces a set of $N$ points that encode the position, movement, and reflectivity of body segments: $\mathbf{p}_i = (x_i, y_i, z_i, v_i, I_i), \quad i = 1, \ldots, N$, where $(x, y, z)$ are the spatial coordinates, $v$ is the Doppler velocity, and $I$ is the signal intensity. The effectiveness of each point cloud attribute is evaluated in our ablation analysis (Sec. V-D).

### B. High-fidelity Deep Inertia Reconstruction Model

Our goal is to directly regress segmental accelerations from point clouds, bypassing explicit skeleton reconstruction and numerical differentiation. Fig 2 illustrates the design of the inertia reconstruction model. The model first extracts spatial features with CNN and squeeze-and-excitation (SE), then uses LSTM to model temporal relationships across frames. The design is detailed below.

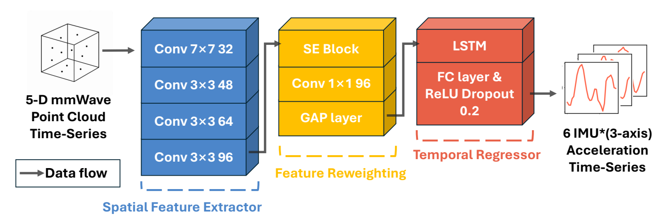


Fig. 2. Design of the high-fidelity inertia reconstruction model.

**Spatial Feature Extractor.** Previous research has shown that convolutional neural networks (CNNs) are well-suited for mmWave radar point cloud-based human pose estimation [8]. Building on this, we first adopt a *CNN backbone* to directly regress segmental accelerations, leveraging its strong ability to extract local spatial features efficiently.

**Feature Reweighting.** We further introduce feature attention via *squeeze-and-excitation (SE) blocks*, which adaptively reweight features based on learned importance to capture the varying importance of spatial coordinates, Doppler velocity, and intensity for acceleration inference. SE blocks have proven effective in point cloud processing, particularly for sparse, heterogeneous data like our $(x, y, z, v, I)$ features [9]. Recent work shows that integrating SE blocks significantly improves recognition accuracy with minimal computational overhead. In our design, the SE block is placed after the fourth convolutional layer, so that attention is applied to semantically rich, high-level features. This placement allows the model to learn more meaningful feature importance.

**Temporal Regressor.** While CNN-SE modules are effective for learning local spatial patterns, they struggle to model the temporal dynamics essential for human motion. To overcome this, we introduce an LSTM module to explicitly capture sequential context over time and reconstruct the final inertia measurements.

This joint spatial-temporal approach allows the model to better track motion trends and transitions, leading to smoother and more realistic acceleration predictions, with improved performance across different subjects and actions.

### C. Kinematics-guide Model Optimization

Optimizing the above model for high-fidelity estimation from mmWave point clouds is fundamentally challenging. Standard regression objectives such as mean squared error (MSE) or mean absolute error (MAE) have significant drawbacks when they are used to optimize our model: MSE is highly sensitive to outliers, while MAE under-penalizes large errors, often resulting in slow convergence and suboptimal accuracy. To address these limitations, we design a composite loss function that combines statistical robustness with biomechanical plausibility, informed by domain knowledge from human Kinematics. Our loss captures three key aspects of physically meaningful results:

- *Magnitude Consistency:* We use the Huber loss for its robustness to outliers and stable convergence, together with an amplitude consistency term that directly matches the vector norms of inertia estimation and ground truth. This approach overcomes the limitations of MSE and MAE and improves reliability under noise.
- *Directional Consistency:* To align estimated and true acceleration vectors, we employ a Pearson correlation loss, which promotes linear agreement in direction, a property often overlooked by standard regression losses.
- *Temporal Smoothness and Minimal Jerk:* Human motion in ADL is typically smooth, with few abrupt changes in acceleration (jerk) [10]. We add the first and second derivatives of the reconstructed inertia to the optimization to ensure natural temporal smoothness and regulate jerk in the estimated inertia measurements.

## IV. Implementation and Benchmark

**System Implementation.** mV-IMU is implemented in Python 3.10 and based on PyTorch 2.0. Training and evaluation were conducted on an desktop PC with Intel i7-14700K CPU, 64 GB RAM, and an NVIDIA RTX 4070 Ti SUPER GPU (16 GB VRAM). For optimization, We used the Adam optimizer with a learning rate of $3 \times 10^{-4}$, batch size 64, and up to 300 epochs.

**Data Preparation** We evaluate mV-IMU using the publicly available mRI dataset [5], which provides synchronized IMU readings and mmWave point clouds from 20 human subjects. We transform both mmWave point clouds and IMU acceleration labels into a common coordinate system using the T-pose calibration data. Frames corrupted by noise are identified and removed through anomaly detection, ensuring high-quality labels.

**Evaluation Metrics** System performance is comprehensively assessed using MAE, RMSE, and Pearson correlation, across six key body segments and a range of representative actions. Two evaluation protocols are adopted: signal-level random splitting (S1) and subject-level splitting (S2). For both protocols, training, validation, and test sets are splitted as 60%-20%-20%. The S2 protocol directly measures generalization to unseen subjects, which is crucial for real-world deployment.

## V. Performance Evaluation

### A. Overall Results

We first demonstrate qualitative results. The mV-IMU readings are visualized against ground truth IMU signals in Fig. 3. Our system captures key kinematic features such as peaks, valleys, and zero crossings across different body parts, demonstrating its ability to recover fine-grained temporal dynamics from mmWave signals. We then performed quantitative evaluation and observe vM-IMU readings achieve an average MAE of $0.1186\,\mathrm{m/s^2}$, RMSE of $0.0121\,\mathrm{m/s^2}$, and a high Pearson correlation coefficient of $0.8169$.

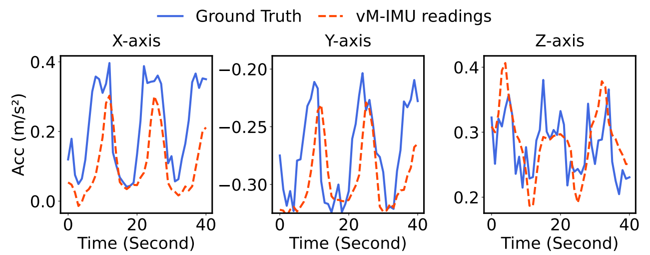


Fig. 3. Three-axis knee acceleration readings from vM-IMU comparing to ground truth readings from a wearable IMU.

Per-IMU error analysis shown in Fig. 4 further reveals that vM-IMU consistently achieve the lowest MAE on pelvis (approximately $0.027\,\mathrm{m/s^2}$, $\rho = 0.83$). The lower limbs also have low MAE. This is because of their large mmWave reflecting area. In contrast, the mV-IMU readings on wrists and head exhibit relatively higher errors, especially along axes with limited radar coverage or frequent occlusions. In addition, vM-IMU generalizes well to unseen subjects. We observe slight fidelity drop for small or rapidly moving extremities, such as the wrists and head. We believe vM-IMU's fidelity on wrist and head can be further improved by cross-body analysis utilizing readings from low-error body segments.

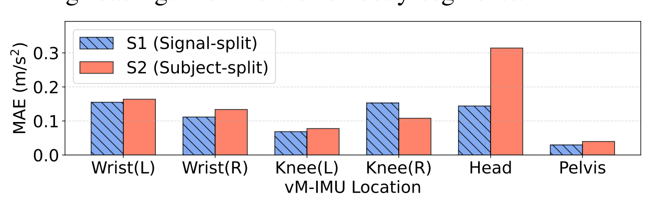


Fig. 4. Mean absolute errors of vM-IMU's readings on different body areas.

### B. How much better is the new paradigm?

As shown in Fig. 5, vM-IMU has consistently and substantially higher signal fidelity compared to the pose-based inertial measurement paradigm. Notably, the pose-based paradigm exhibits large errors and instability at distal joints such as the wrists, where noise amplification and signal occlusion are most severe. These results confirm that the proposed direct inertial measurement paradigm outperforms the traditional pose-based differentiation.

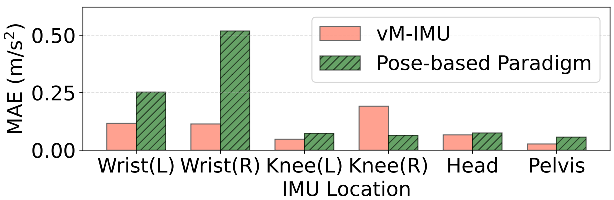


Fig. 5. Mean absolute error of vM-IMU readings and pose-based estimation.

*C. Case Study: Gait Analysis*

Gait segmentation is a critical pre-processing step used in most IMU-based gait analysis frameworks. We conducted this case study using gait segmentation to answer the question: to what extent the high-fidelity vM-IMU readings can bridge the existing wisdom on wearable IMUs? Specifically, gait segmentation is conducted on ground-truth, pose-based estimation, and vM-IMU readings. We then calculate the length error of the segments, which are reported in Fig. 6. We observe that **91.7%** of gait segments from vM-IMU have errors less than 70 ms (median = 40 ms), whereas nearly half of the segments from pose-based method have larger errors.

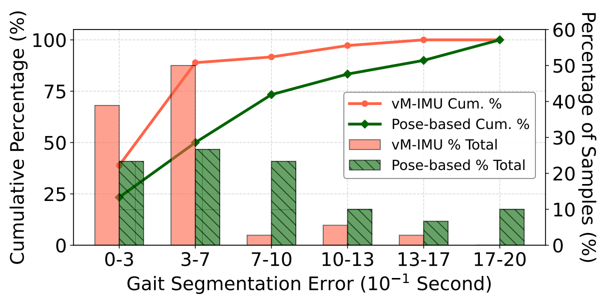


Fig. 6. Gait segmentation error based on vM-IMU and pose-based method.

*D. System Ablation Analysis*

We conducted systematic ablation studies to analyze the importance of the proposed Feature Reweighting and Temporal Regressor (FRTR). As reported in Fig. 7 (left), FRTR effectively reduces measurement errors. Moving forward, we studied the impact of mmWave inertial features on vM-IMU's fidelity, which is shown in Fig. 7 (right). Using all three feature (i.e., spatial coordinates, Doppler velocity, and intensity) yields the highest fidelity. Without either Doppler or intensity, the error increases significantly. This highlights the importance of these two channels for robust and generalizable results.

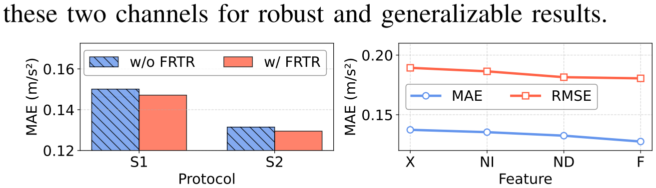


Fig. 7. Ablation study. Left: MAE w/ and w/o Feature Reweighting and Temporal Regressor (FRTR). Right: MAE and RMSE with various features.

## VI. Conclusion and Future Work

This work addresses the limitations of pose-based inertial estimation in device-free sensing by introducing mV-IMU, a novel mmWave-enabled virtual inertial measurement framework. By directly reconstructing body accelerations from mmWave signals, without relying on pose tracking or numerical differentiation, mV-IMU avoids error amplification and instability inherent in prior approaches. Our results demonstrate that mmWave sensing, when coupled with deep learning and kinematics-guided optimization, can deliver high-fidelity inertial measurements comparable to wearable IMUs, enabling a practical and privacy-preserving solution for continuous, non-intrusive human activity monitoring.

To further advance the mmWave-enabled Virtual IMU paradigm, several extensions are worth pursuing. First, comprehensive studies on patients who can benefit from the proposed new paradigm can be conducted to quantify the effectiveness of vM-IMU on helping clinical decision-making. Second, new algorithms considering the Kinematics relation among body areas can be introduced to better mitigate the impact of mmWave occlusion. Lastly, real-time on-device inertia computing can be explored to preserve more user privacy.

## References


[1] A. Favata, R. Gallart-Agut, R. Pàmies-Vilà, C. Torras, and J. M. Font-Llagunes, "Imu-based systems for upper-limb kinematic analysis in clinical applications: A systematic review," *IEEE Sensors Journal*, vol. 24, no. 18, pp. 28 576–28 594, 2024.

[2] H. Kwon, C. Tong, H. Haresamudram, Y. Gao, G. D. Abowd, N. D. Lane, and T. Ploetz, "Imutube: Automatic extraction of virtual on-body accelerometry from video for human activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–29, 2020.

[3] Y. Ren, Z. Wang, Y. Wang, S. Tan, Y. Chen, and J. Yang, "Gopose: 3d human pose estimation using wifi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 2, Jul. 2022. [Online]. Available: https://doi.org/10.1145/3534605

[4] H. Xue, Q. Cao, C. Miao, Y. Ju, H. Hu, A. Zhang, and L. Su, "Towards generalized mmwave-based human pose estimation through signal augmentation," in *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, 2023, pp. 1–15.

[5] S. An, Y. Li, and U. Ogras, "mri: multi-modal 3d human pose estimation dataset using mmwave, rgb-d, and inertial sensors," in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, ser. NIPS '22. Red Hook, NY, USA: Curran Associates Inc., 2022.

[6] J. Zhang, R. Xi, Y. He, Y. Sun, X. Guo, W. Wang, X. Na, Y. Liu, Z. Shi, and T. Gu, "A survey of mmwave-based human sensing: Technology, platforms and applications," *Commun. Surveys Tuts.*, vol. 25, no. 4, p. 2052–2087, Oct. 2023. [Online]. Available: https://doi.org/10.1109/COMST.2023.3298300

[7] S. An and U. Y. Ogras, "Mars: mmwave-based assistive rehabilitation system for smart healthcare," *ACM Trans. Embed. Comput. Syst.*, vol. 20, no. 5s, Sep. 2021. [Online]. Available: https://doi.org/10.1145/3477003

[8] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-pose: Real-time human skeletal posture estimation using mmwave radars and cnns," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10 032–10 044, 2020.

[9] Z. Jing, H. Guan, P. Zhao, D. Li, Y. Yu, Y. Zang, H. Wang, and J. Li, "Multispectral lidar point cloud classification using sepointnet++," *Remote Sensing*, vol. 13, no. 13, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/13/2516

[10] C. Wang, L. Peng, Z.-G. Hou, L. Luo, S. Chen, and W. Wang, "Experimental validation of minimum-jerk principle in physical human-robot interaction," in *International Conference on Neural Information Processing*. Springer, 2018, pp. 499–509.