Sensing Life in Stillness: Unified Dynamic and Static Human Mesh Reconstruction with mmWave Radar

LIN CHEN, The Hong Kong University of Science and Technology (Guangzhou), China
CONG LI, The Hong Kong University of Science and Technology (Guangzhou), China
SHUXIN ZHONG*, The Hong Kong University of Science and Technology (Guangzhou), China
JUN CHEN, The Hong Kong University of Science and Technology (Guangzhou), China
YUFEI WEN, The Hong Kong University of Science and Technology (Guangzhou), China
HAOTIAN SONG, The Hong Kong University of Science and Technology (Guangzhou), China
KAISHUN WU*, The Hong Kong University of Science and Technology (Guangzhou), China

Continuous rehabilitation monitoring outside clinics is critical for long-term recovery, yet existing modalities fall short. Vision- and wearable-based systems raise privacy and compliance concerns, while RF-based sensing, despite contactless, is fundamentally *motion-dependent*—blind to the stillness that characterizes balance, endurance, and postural control. We observe that conventional static clutter removal not only suppresses environmental reflections but also erases the body's involuntary micro-motions, such as breathing, heartbeat, and subtle sway. Our key insight is that these micro-motions are not noise but information, encoding physiological vitality even in apparent stillness. Realizing this shift—from detecting motion to perceiving life within stillness—introduces two fundamental challenges: (**C1**) the echoes of these micro-motions are orders of magnitude weaker than static clutter, spectrally overlap within the near-zero Doppler region and spatially co-located with dominant reflections; and (**C2**) mmWave reflections are inherently sparse and geometry-agnostic, lacking the structural priors required to recover body's shape and pose across users and environments. To address these, we design `mmRehab`, a transformative mmWave sensing system for rehabilitation, extending radar perception beyond motion to enable physiological interpretation even when users remain still. Within `mmRehab`, *Micro-motion Feature Extraction* addresses **C1** through beamforming-based spatial isolation and micro-Doppler temporal discrimination, amplifying respiration- and posture-related cues; *Geometry-aware Knowledge Transfer* addresses **C2** via depth-guided distillation, transferring structural priors from vision to radar representations for robust generalization. Extensive experiments on both dynamic and static rehabilitation tasks show that `mmRehab` reduces 3D reconstruction errors by over 24% and generalizes robustly to unseen users, distances, and orientations—demonstrating the feasibility of unified radar perception for motion and micro-motion rehabilitation monitoring.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: wireless sensing, mmwave, 3D human mesh, rehabilitation

---

*Corresponding authors.

Authors' Contact Information: Lin Chen, The Hong Kong University of Science and Technology (Guangzhou), China, lchen297@connect.hkust-gz.edu.cn; Cong Li, The Hong Kong University of Science and Technology (Guangzhou), China, cli166@connect.hkust-gz.edu.cn; Shuxin Zhong, The Hong Kong University of Science and Technology (Guangzhou), China, shuxinzhong@hkust-gz.edu.cn; Jun Chen, The Hong Kong University of Science and Technology (Guangzhou), China, jchen512@connect.hkust-gz.edu.cn; Yufei Wen, The Hong Kong University of Science and Technology (Guangzhou), China, ywen196@connect.hkust-gz.edu.cn; Haotian Song, The Hong Kong University of Science and Technology (Guangzhou), China, hsong425@connect.hkust-gz.edu.cn; Kaishun Wu, The Hong Kong University of Science and Technology (Guangzhou), China, wuks@hkust-gz.edu.cn.

ACM 2474-9567/2026/3-ART1
https://doi.org/10.1145/3790117

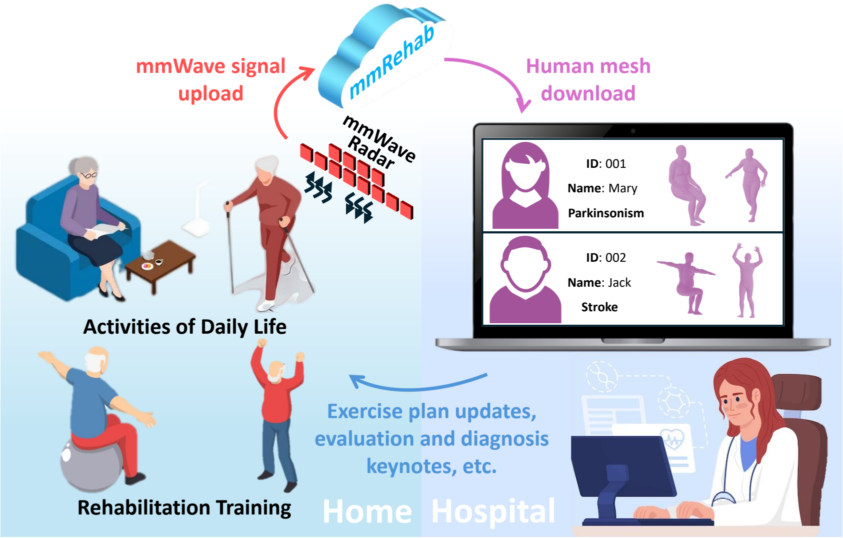


Fig. 1. mmRehab turns everyday living spaces into continuous rehabilitation environments—no wearables, no cameras, just wireless reflections. By reconstructing human meshes from mmWave signals, it allows clinicians to remotely observe postures, update training plans, and maintain 24/7 continuity of care across home and hospital.

ACM Reference Format:
Lin Chen, Cong Li, Shuxin Zhong, Jun Chen, Yufei Wen, Haotian Song, and Kaishun Wu. 2026. Sensing Life in Stillness: Unified Dynamic and Static Human Mesh Reconstruction with mmWave Radar. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 1, Article 1 (March 2026), 25 pages. https://doi.org/10.1145/3790117

1 INTRODUCTION

Over 2.4 billion people live with conditions that demand continuous rehabilitation—a number expected to surge as longer lifespans meet rising chronic disease and disability [74]. Yet rehabilitation remains constrained by scarce clinical resources and limited accessibility: most patients cannot be monitored once they leave the hospital, causing lapses that undo weeks of recovery [35, 61]. Bridging this gap requires more than **remote supervision**—it demands *high-fidelity sensing systems* that can unobtrusively capture human motion and physiology across daily environments [20, 44] (as illustrated in Fig. 1).

However, current sensing modalities fall short of this vision. Vision-based systems [25, 36, 38, 39, 45, 56, 77, 90] achieve remarkable accuracy but collapse under occlusions, lighting variation, and privacy concerns. Wearables [2, 6, 30, 55, 59] mitigate these issues yet rely on user compliance and continuous maintenance. Environment-embedded sensors such as tactile carpets [13] offer unobtrusive feedback but remains spatially confined. Recently, RF-based sensing [4, 15, 34, 78–80, 86, 87]) has emerged as a compelling alternative, promising contactless and privacy-preserving perception. Yet despite these advances, RF sensing suffers from a unique limitation: it perceives only **motion**, not **control**. When the body stop moving, these system fail silent—blind to the stability, endurance, and subtle balance that define true recovery [24, 64].

Consider a simple bridge exercise [17, 21]: once the posture is held, the patient appears perfectly motionless. Yet beneath this stillness, the body is far from static—the redistribution of pressure—from the lumbar region toward the scapulae and heels encodes postural stability, muscle engagement, and recovery progress. Capturing such **life within stillness** requires a paradigm shift: enabling mmWave radar not only to detect movement, but to perceive the stability that underlies it.

Operating at tens of gigahertz, mmWave radar is exquisitely sensitive to these displacements [57]. Yet the root cause of existing limitation lies in how radar interprets its own signals. Conventional pipelines rely on *static clutter removal (SCR)* [80] to suppress zero-Doppler reflections from stationary objects, inadvertently erasing weak echoes from seemingly still users. The human body is never truly motionless: even in "static" postures, involuntary activities such as breathing, heartbeat, postural sway produce millimeter-scale **micro-motions** [10]. These faint vibrations are **not noise, but physiology**—subtle signatures that encode both **presence and posture** in apparent stillness [84]. This observation opens a new opportunity: to redefine radar perception from detecting motion to perceiving **life within stillness**. [1]

Realizing this vision, however, introduces two fundamental challenges: **(C1) Extracting micro-motions under overwhelming clutter:** life-induced reflections are orders of magnitude weaker than static returns from walls and furniture [81], and they occupy the same near-zero Doppler band, making them nearly indistinguishable in both frequency and angle domains. **(C2) Reconstructing human geometry from sparse radar echoes:** even when preserved, radar reflections are inherently view-dependent and geometry-agnostic [11]—comprising discrete scattering points without explicit cues of body topology or joint kinematics [67]. Consequently, inferring a coherent 3D human form from these echoes becomes a highly ill-posed inverse problem, especially under quasi-static conditions [91].

To addess these challenges, we design `mmRehab`, a mmWave sensing system tailored for rehabilitation, transforming radar from a motion detector into a **micro-motion observer** capable of reconstructing fine-grained 3D meshes even when users appear completely still. Specifically, to address **C1**, *Micro-motion Feature Extraction* couples adaptive beamforming with micro-Doppler temporal analysis to isolate life-induced oscillations while suppressing static clutter; and to address **C2**, *Geometry-aware Knowledge Transfer* leverages **depth maps** [73] as an intermediate geometric bridge to transfer structural priors from vision-based models, enabling radar-only 3D human mesh reconstruction with semantic fidelity and cross-user generalization. In summary, our contributions in this paper are:

- `mmRehab` redefines radar perception by inverting a long-standing assumption: *what was once suppressed as "clutter" is now recognized as a signal of living stillness.* We reveal that the zero-Doppler reflections, traditionally discarded by static clutter removal, encode millimeter-scale physiological dynamics—breathing, heartbeat, and postural sway—enabling radar to perceive the quiet dynamics that sustain human stability and recovery.
- We realize this vision through `mmRehab`, a transformative mmWave sensing system that reconstructs fine-grained 3D meshes even under stillness. At its core, `mmRehab` introduces two technical designs: (i) *Micro-motion Feature Extraction* addresses **C1** by coupling adaptive beamforming with micro-Doppler temporal analysis to isolate life-induced micro-signals from dominant static clutter; and (ii) *Geometry-aware Knowledge Transfer* addresses **C2** by leveraging depth maps as a geometric bridge to transfer structural priors from large-scale visual datasets to endow radar with geometric awareness.
- We conduct extensive experiments under both dynamic and static rehabilitation scenarios. `mmRehab` reduces the average Vertex Error (V) from 7.21 cm → 4.75 cm and the Joint Localization Error (S) from 6.44 cm → 3.38 cm for static postures, achieving similar gains for dynamic motions (V: 6.92 cm → 4.12 cm, S: 5.87 cm → 3.25 cm). Even on unseen users, `mmRehab` maintains high accuracy (V: 6.77 cm, S: 5.97 cm for dynamic; V: 7.46 cm, S: 5.74 cm for static) and remain robust under varied distances (2-4 m) and orientations (up to 60 °), confirming its real-world viability for in-home rehabilitation monitoring.

---

[1] Technically, this "life within stillness" arises from **micro-motions**—minute, involuntary oscillations such as breathing and postural sway that reveal physiological vitality even in static postures.

To support the research community, we make our source code publicly available at https://github.com/leenchen0/mmRehab.

2 RELATED WORKS

This section reviews related work from two complementary perspectives: (1) **rehabilitation monitoring**, which examines sensing modalities used for assessing patient motion and recovery, and (2) **wireless human mesh reconstruction**, which extends wireless perception from coarse activity recognition to fine-grained geometric modeling of the human body.

2.1 Rehabilitation Monitoring

Rehabilitation monitoring traditionally relied on vision-based [22, 47, 55, 83, 89] or wearable systems [45, 56, 90]. While effective in controlled settings, these methods remain *intrusive and non-scalable*—cameras compromise privacy and require line-of-sight, whereas wearables depend on patient compliance. To overcome these limitations, recent work explores *wireless sensing* as a non-contact, privacy-preserving alternative [3, 4, 27, 43, 68, 75]. These efforts fall into two main directions: (1) *Signal-based activity recognition* [27, 68], which classifies rehabilitation motions directly from Doppler or amplitude patterns; and (2) *Pose-based motion estimation* [3, 4, 43, 75], which reconstructs skeletal keypoints for fine-grained motion analysis. However, motion tracking alone cannot comprehensively explain recovery [70]; rehabilitation depends on the *biomechanics behind stillness*—how balance, load, and subtle corrections sustain the body [24]. Capturing these nuances requires *continuous 3D mesh reconstruction*, which models the full body surface to reveal the invisible dynamics of posture and pressure [49].

Significance of Continuous Mesh Reconstruction for Rehabilitation Monitoring. The ability to continuously reconstruct full-body 3D meshes provides a high-resolution geometric foundation for analyzing body-surface biomechanics [52]. This mesh-level representation supports diverse rehabilitation analyses, such as monitoring pressure redistribution, identifying improper load transfer, and evaluating postural balance during recovery [42, 66]. For instance, in bridge exercises [17, 21], the reconstructed meshes provide geometric cues to evaluate whether the stress is likely shifted toward the scapulae and heels rather than the lumbar region. As clinical guidance emphasizes that the need for continuous, 24/7 posture and pressure monitoring [41, 82], continuous 3D mesh reconstruction offers a clinically meaningful foundation for long-term pressure ulcer risk assessment. Recent work has demonstrated that recovering the precise 3D body mesh is essential for accurately inferring contact pressure distributions and localizing peak pressure points on the body surface [16], enabling the capture of evolving body-support interactions.

2.2 Wireless Human Mesh Reconstruction

Therefore, recent works have begun to enable wireless-based human mesh reconstruction [78–80], extending human sensing from sparse skeletons toward full surface geometry. Hybrid methods (MI-Mesh [19], Immfusion [12]) leverage complementary modalities to reconstruct detailed human meshes. Pure RF systems [1, 8, 40, 62, 63, 79, 87, 88] are also emerging: for example, mmMesh [80] uses commercial mmWave radar to build dynamic human meshes from point-cloud data, and M^{4}sh extends this to multi-subject 3D mesh reconstruction [78]. Other wireless-signal based systems reconstruct skeletons or coarse meshes using WiFi [23, 34, 58, 71, 72].

Despite these advances, existing systems remain fundamentally **motion-centric**—they rely on Doppler variations and thus fail to perceive users who appear still. Yet, static postures encode critical cues for rehabilitation, such as seated balance training [7], where micro-motions (e.g., breathing, postural sway) reflect balance, endurance, and recovery stability. `mmRehab` overcomes this limitation by reinterpreting radar reflections—not as mere motion indicators, but as **signatures of life within stillness**. This paradigm shift enables accurate reconstruction of both dynamic and static postures, supporting continuous, contactless rehabilitation assessment.

3 PRELIMINARIES

To enable human mesh reconstruction from radar echoes, we first review the key sensing and modeling foundations of our system: mmWave radar for physical signal acquisition and Skinned Multi-Person Linear (SMPL) model for human body representation.

3.1 Mmwave Sensing

Our system employs a commercial off-the-shelf (COTS) mmWave radar that continuously scans the environment and captures reflections from all *entities* [2] within the sensing area. The radar operates using frequency-modulated continuous wave (FMCW) signaling, wher the transmitter (TX) emits a sequence of chirps whose frequency increases linearly over a bandwidth B within a chirp duration T_c. The reflected signal received by the receiver (RX) is mixed with the transmitted waveform to produce an intermediate-frequency (IF) beat signal:

$$x_{IF}(t) = A_r \, e^{j2\pi\cdot\tau(r,c)} \left[f_0 + \frac{B}{T_c} t - \frac{B}{2T_c} \tau(r,c) \right], \tag{1}$$

where f_0 denotes the start frequency, A_r represents an amplitude after propagation attenuation, and $\tau(r,c)$ is the round-trip delay determined by range r and signal speed c. The IF signal encodes the spatial and temporal structure of reflections from all entities, serving as the raw measurement for reconstructing human geometry in subsequent stages.

3.2 Human Body Representation

To recover anatomically consistent human structures from radar-derived features, we adopt the SMPL model [49]—a widely used parametric representation that disentangles human variation into shape and pose components. The shape parameters $\boldsymbol{\beta} \in \mathbb{R}^{10}$ captures identity-dependent attributes such as body height and weight, while the pose parameters $\boldsymbol{\theta} \in \mathbb{R}^{72}$ encode global orientation and 23 joint rotations. Using linear blend skinning, SMPL generates a triangulated mesh with $N = 6890$ vertices and $K = 24$ joints. Its compact 82-parameter design provides a differential, low-dimensional representation that enables efficient and high-fidelity reconstruction from sparse radar signals.

4 DESIGN OF MMREHAB

This section begins with an overview of the system pipeline and its core design principles, followed by detailed descriptions of each component.

4.1 System Overview

Fig. 2 presents `mmRehab`, a mmWave sensing system tailored for rehabilitation that redefines how radar perceives human presence—not by tracking motion, but by sensing physiological micro-motions that reveal life within stillness. The system comprises four synergistic components that together reconstruct full-body 3D human meshes from commodity mmWave radar under both dynamic and quasi-static rehabilitation scenarios.

- **Geometry-aligned Data Generation (Sec. 4.2).** We first build a multimodal dataset pairing real and simulated radar signals with synthetic depth maps rendered from 3D human meshes, creating *geometry-consistent supervision* that aligns radar and visual semantics in a shared spatial frame.
- **Hierarchical Physiological Feature Extraction (Sec. 4.3).** The radar stream is then processed through a *Dynamic Spatial-Kinematic Mapping* and a *Static Vital Micro-motion Extraction*, jointly capturing large-scale body movement and subtle physiological oscillations such as respiration and postural sway—enabling radar to perceive both motion in space and vitality in stillness (addresses **C1**).

---

[2] An *Entity* refers to any detected target, categorized as a *User* or an *Object*.

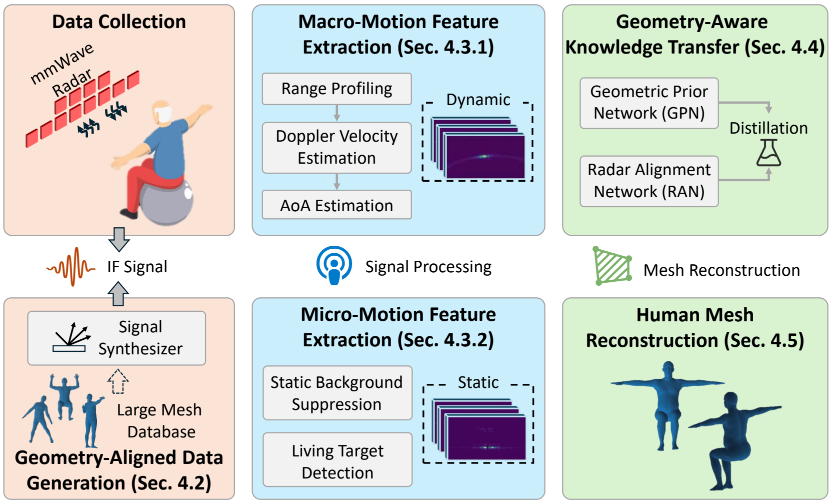


Fig. 2. System framework of mmRehab.

- **Geometry-aware Knowledge Transfer (Sec. 4.4).** To bridge sparse radar observations and dense human geometry, depth maps serve as an intermediate geometric modality—sharing spatial structure with 3D meshes while remaining compatible with radar's range domain—allowing a vision-based teacher to distill structural priors into a radar-based student (addresses **C2**). This enables meaningful and generalizable perception across unseen users and environments.
- **Human Mesh Reconstruction (Sec. 4.5).** Finally, we decode the refined radar features into full-body SMPL parameters—pose, shape, and translation—using a differentiable rotation representation that guarantees anatomical continuity and stable mesh reconstruction.

4.2 Geometry-aligned Data Generation

To establish cross-modal supervision between radar and geometry, we construct paired datasets of depth maps and simulated radar signals derived from AMASS dataset [50]. AMASS provides over 2,400 minutes of motion data covering diverse human poses and shapes, from which we obtain 3D mesh sequences parameterized by SMPL.

A virtual camera is positioned at $(0, -2, 0)$ m with a resolution of 64×64, and the Hidden Point Removal (HPR) algorithm [37] is applied to project visible vertices:

$$x_i = \min(W, \max(1, \text{round}(x \cdot \text{scale}))), \quad y_i = \min(H, \max(1, \text{round}(z \cdot \text{scale}))). \tag{2}$$

Here, (x, z) denote 3D vertex coordinates, (x_i, y_i) are projected 2D locations, and scale normalizes the 3D space to image coordinates ($W = H = 64$). For each visible pixel (x_i, y_i), the depth is defined as the Y-axis distance from the camera to the corresponding vertex:

$$d(x_i, y_i) = |y_{\text{camera}} - y_{\text{vertex}}|. \tag{3}$$

The view-consistent depth maps preserve pose-dependent body geometry and serve as the teacher's supervision signal (Fig. 3).

In parallel, we synthesize radar data from the same mesh sequences following the RF-Genesis framework [15], which simulates intermediate-frequency reflections based on mesh geometry and material properties. To ensure diversity, radar positions and orientations are randomized across samples—placed at a height of 0.8 m, with the horizontal distance uniformly sampled between 2 m and 6 m, and the azimuth angle varied from 0° to 90°. The generated IF signals are converted into radar spectra and used alongside real radar recordings during training.

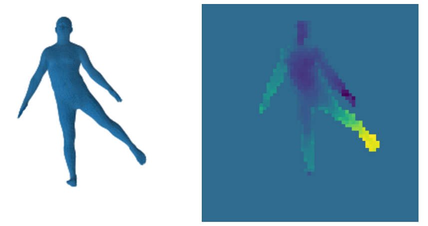

(a) 3D SMPL mesh (b) Depth image rendered from the mesh via virtual projection

Fig. 3. Depth image generation from 3D mesh. A virtual camera projects visible mesh vertices from the SMPL model (a) onto a 2D plane to produce a rendered depth image (b), where pixel intensity encodes distance to the camera. The synthetic depth maps serve as geometric intermediates between vision-based meshes and radar reflections during knowledge distillation.

4.3 Hierarchical Physiological Feature Extraction

Existing mmWave sensing systems primarily focus on *dynamic body movements*, such as walking or turning, where large-scale range and Doppler shifts provide strong geometric cues [85]. While effective for motion tracking, these methods remain blind to **quasi-static physiological activities**—minute oscillations from respiration or subtle posture adjustments—that lie below the resolution of conventional Doppler processing.

To overcome this limitation, `mmRehab` introduces *Hierarchical Physiological Feature Extraction* component that extends mmWave perception from motion geometry to life-state awareness. The first branch, *Macro-motion Feature Extraction* (Sec.4.3.1), reconstructs large-scale body motions—a well-understood regime in mmWave sensing [86]—provide spatial context for subsequent physiological inference. In contrast, our second branch, *Micro-motion Feature Extraction* (Sec.4.3.2), targets the previously unmodeled sub-centimeter physiological oscillations that conventional Doppler processing fails to capture.

4.3.1 Macro-motion Feature Extraction. We construct a spatial–kinematic map for dynamic users, capturing their location, motion, and direction directly from the raw IF signals.

Range Profiling. Since the IF beat frequency f scales linearly with target range r as:

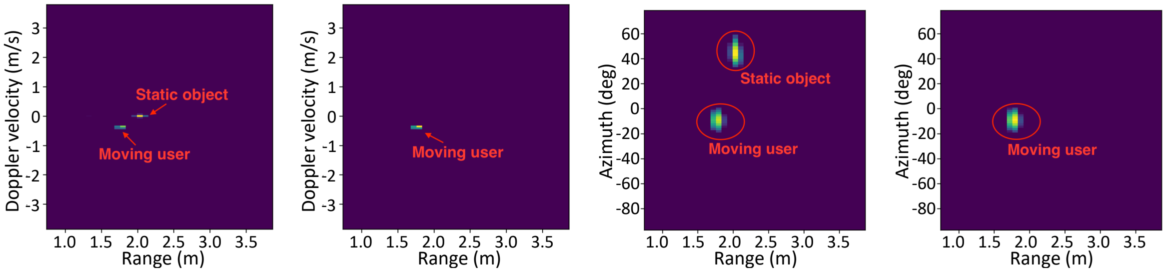
$$r = \frac{cfT_c}{2B}, \tag{4}$$

a fast-time FFT (Range-FFT) yields the range profile of all reflectors in the scene.

Static Clutter Removal (SCR). Range-domain profiles contain both static and dynamic components. To suppress static structures (walls, furniture, etc.) that otherwise form a strong near-zero Doppler band, we apply SCR—subtracting the slow-time mean per range bin as in [79].

Doppler Velocity Estimation. The target's radial velocity is inferred from the slow-time phase evolution ω across chirps:

$$v = \frac{\lambda}{4\pi T_c}\,\omega. \tag{5}$$



(a) RDM w/o SCR (b) RDM w/ SCR (c) RAM w/o SCR (d) RAM w/ SCR

Fig. 4. Effect of static clutter removal (SCR) on radar feature maps. (a-b) Range-Doppler maps (RDMs) and (c-d) Range-Azimuth maps (RAMs) with or without SCR. Without SCR, strong zero-Doppler reflections from static objects obscure user motion, while applying SCR effectively suppresses these unwanted signals—highlighting dynamic targets and improving downstream macro-motion analysis.

An additional FFT along slow time (Doppler-FFT) forms the *Range–Doppler Map (RDM)*, which encodes target speeds along the radial dimension.

Angle-of-Arrival Estimation (AoA). Range or Doppler alone cannot separate multiple entities at the same distance. To resolve spatial ambiguity, `mmRehab` employs a MIMO antenna array that estimates azimuth and elevation angles. This is achieved by correlating the array snapshots with candidate steering vectors in the angular domain:

$$x^{(N_a \times N_e)} = \sum_{m=1}^{M} \alpha_m \, s\big(\theta_{(a,e)m}\big) + n, \tag{6}$$

where M is the number of incoming signals, α_m are their complex amplitudes, and $s(\cdot)$ denotes the steering vector defined by inter-element phase shifts $\Delta\phi_a$ and $\Delta\phi_e$:

$$s(\theta_{(a,e)}, N_a, N_e) = \begin{bmatrix} 1 & \cdots & e^{j(N_a-1)\Delta\phi_a} \\ e^{j\Delta\phi_e} & \cdots & e^{j(\Delta\phi_e+(N_a-1)\Delta\phi_a)} \\ \vdots & \ddots & \vdots \\ e^{j(N_e-1)\Delta\phi_e} & \cdots & e^{j((N_e-1)\Delta\phi_e+(N_a-1)\Delta\phi_a)} \end{bmatrix}, \tag{7}$$

where $\Delta\phi_a$ and $\Delta\phi_e$ denote the inter-element phase increments in azimuth and elevation, respectively. By correlating x with steering vectors over candidate angles—implemented efficiently as a 2D FFT across receiver elements—we obtain the Angle Profile (i.e., *Range-Elevation Map (REM)* and *Range-Azimuth Map (RAM)*), revealing each reflector's direction.

Spatial–Kinematic Signature. Combining the RDM with the Angle Profile yields a complete spatial–kinematic signature, capturing range, velocity, and direction for all entities.

Visualization of SCR Effects on RD and RA Maps. To visualize the effect of SCR, we collect radar data in an empty room containing a single chair, while a volunteer moves within the sensing area. This setup enables a clear comparison of how SCR alters both the RDM and the RAM, as shown in Fig. 4. Without SCR, the RDM exhibits a strong near-zero Doppler band originating from the static chair, which leaks into dynamic bins and masks the user's motion signatures (Fig. 4(a)). After applying SCR, these stationary components are effectively suppressed, and the moving subject appears as a distinct energy ridge along the corresponding range–velocity cells (Fig. 4(b)). A similar pattern is observed in the angular domain. Before SCR, static reflections from the chair generate broad azimuthal sidelobes that obscure the true direction of the moving target (Fig. 4(c)). After SCR, dynamic energy becomes localized around the volunteer's bearing, resulting in a cleaner angular profile (Fig. 4(d)). Overall, SCR substantially enhances spatial-kinematic clarity by removing near-zero Doppler clutter, isolating dynamic motion signatures, and improving downstream macro-motion analysis.

4.3.2 Micro-motion Feature Extraction. `mmRehab` designs *Micro-motion Feature Extraction* to suppress background clutter and retain faint physiological micro-motions that indicate life within stillness. This component operates in two stages: (1) *Static Background Suppression* spatially disentangles static reflections via adaptive beamforming, and (2) *Living Target Detection* temporally isolates micro-motion components from the residual clutter—together enabling detection of living users even in fully still scenes.

Static Background Suppression. In static scenario, each detected *Entity*—either a *User* or an *Object*—contributes to the radar echoes within static frames. These echoes are a superposition of strong, stable reflections from environmental *Objects* (walls, floors, furniture) and weaker returns from static *Users* (sitting, standing, or supine). Our key insight is that static clutter, traditionally treated as interference [5, 80], actually encodes the scene's spatial priors—including subtle scattering from stationary users. By statistically modeling these near-zero Dopplers reflections, we transform clutter from noise into signal, enabling recovery of the spatial profile of motionless human bodies. However, when strong *Object*-level clutter (e.g., walls or furniture) coexists with weak *User*-level echoes, traditional FFT-based delay-and-sum [69] or Bartlett fixed-weight beamforming [60] suffers from limited angular resolution and dynamic range, where sidelobes and energy leakage easily obscure the weak *User* reflections [69].

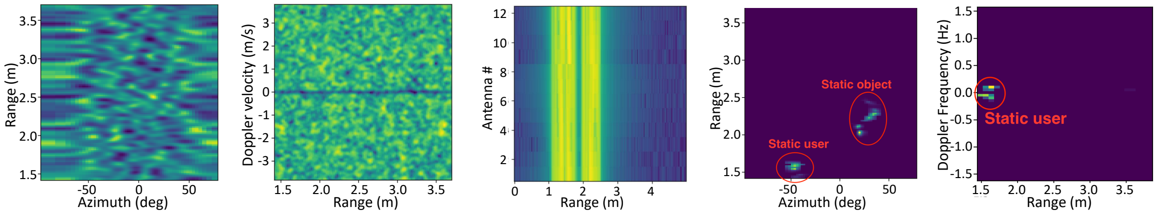
To suppress dominant clutter and enhance weak user reflections, we adopt MVDR (Minimum Variance Distortionless Response) beamformer [9]. MVDR minimizes the output power while enforcing a unity-gain constraint in the steering direction, thereby adaptively nulling dominant interference. This adaptively substantially improves angular contrast and dynamic range even with limited array elements and snapshots, making it particularly effective for detecting weak *User*-level echoes under strong *Object*-level clutter [9, 65]. Unlike conventional beamformers, MVDR requires no explicit target priors; it relies solely on the *scene covariance statistics* to suppress static backgrounds that exhibit spatial signatures distinct from humans [46].

$$P_{MVDR}(r, \theta_{(a,e)}) = \frac{1}{s^H(\theta_{(a,e)}, N_a, N_e) R_x s(\theta_{(a,e)}, N_a, N_e)}, \tag{8}$$

where $R_x = E\{x_{sc}^H x_{sc}\} \approx \frac{1}{N}\sum_{t=1}^{N} x_{sc}^H(t) x_{sc}(t)$ represents the covariance matrix of the static clutter x_{sc}. The resulting 3-D static map is normalized and projected onto 2-D range-angle planes by averaging along the complementary angle dimension, which suppresses isolated spikes while preserving the overall static energy distribution. This yields a *Static Range–Azimuth Map (SRAM)* or *Static Range–Elevation Map (SREM)* that spatially separates multiple static entities (Fig. 5(d)), though identifying which reflection corresponds to a human user remains challenging—a problem we address next.

Living Target Detection. Although standard RDM effectively capture dynamic motions [79], they fail to resolve the faint micro-movements exhibited by stationary users. Typical respiration produces chest displacements of 4–12 mm at 0.2–0.34 Hz [18, 48], corresponding to a Doppler velocity of roughly on ≈ 1.8 cm/s. In contrast, a 77 GHz radar with a 255 μs chirp interval and 128 chirps per frame achieves merely 5.9 cm/s velocity resolution—insufficient to distinguish such subtle variations [51]. Consequently, these physiological cues vanish after SCR, as illustrated in Fig. 4(b) and Fig. 4(d).

Intuitively, one might enhance Doppler sensitivity by tuning radar front-end parameters (e.g., extending inter-chirp intervals or increasing chirp counts [33]), yet such adjustments introduce fundamental trade-offs. Longer inter-chirp intervals reduce the maximum unambiguous velocity, while additional chirps enlarge the radar cube, straining DSP memory and real-time throughput [33]. Aggregating chirps across frames can extend observation windows but disrupts temporal uniformity due to inter-frame blanking for calibration and thermal control,



(a) RAM (b) RDM (c) Static clutter (d) SRAM (e) RMDM

Fig. 5. Detecting static users under still scenes. (a-b) show the raw range-azimuth map (RAM) and range-Doppler map (RAM), while (c) visualizes the strong near-zero Doppler reflections constituting static clutter. Through beamforming and micro-motion analysis, static targets are effectively separated into (d) the static RA map (SRAM) and (e) the range–micro-Doppler map (RMDM), enabling precise identification of static users and surrounding objects.

producing spectral artifacts. These constraints render hardware-level adjustments impractical for continuous, resource-constrained rehabilitation monitoring.

To this end, we introduce a new spectral feature—**Range–MDoppler Map (RMDM)**—that captures micro-motion dynamics invisible to conventional Doppler processing. *The key idea is to apply a temporal FFT over static clutter accumulated across frames, transforming imperceptible sub-centimeter displacements into measurable frequency-domain energy.*

Formally, we first transform the time-domain signal into the range domain using a range-FFT [5]. The received signal at range r, antenna a, and chirp c in frame n can be expressed as:

$$H_{r,a,c} = S_{r,a,c} + D_{r,a,c}(n) + N_{r,a,c}, \tag{9}$$

where $S_{r,a,c}$ denotes static components, $D_{r,a,c}(n)$ represents micro-motion-induced dynamics, and $N_{r,a,c}$ is noise. To isolate the faint physiological component buried in static reflections, we remove the mean clutter $\bar{H}_{r,a}$ via standard SCR and assume the micro-motion remains approximately constant within each short 50 ms frame. We then aggregate n_f consecutive frames (e.g., $n_f = 64$), subtract their temporal mean, and perform an FFT across frames—termed the **micro-Doppler FFT (MDoppler-FFT)**—to reveal slow periodic motion in the frequency domain:

$$\bar{H}_{r,a} = S_{r,a,i} + D_{r,a,i}(n) + \bar{N}_{r,a}, i \in [1, N_c]. \tag{10}$$

Summing across antennas yields the **Range–MDoppler Map (RMDM)**, which amplifies sub-centimeter movements such as breathing or postural sway. As shown in Fig. 5(e), this spectrum exposes a concentrated energy band near 1.5 m, corresponding to the user's position—signals that conventional range–Doppler maps suppress during clutter removal.

Visualization of Static User Detection. To demonstrate mmRehab's ability to separate static users from stationary background clutter, Fig. 5 visualizes how radar signal evolves across successive processing stages. The experiment is conducted in a stationary environment containing a static user and a surrounding object. As shown in Fig. 5(a) and Fig. 5(b), the unprocessed RAM and RDM contain strong clutter energy distributed across azimuth and velocity dimensions, making it difficult to distinguish true static targets from background reflections. By integrating Doppler responses near zero frequency, the dominant scatters, such as walls and furniture, are extracted to form the static clutter field, as shown in Fig. 5(c), which in turn obscures weak user signals. After *Static Background Suppression*, the static RA map (SRAM) clearly shows the static user and surrounding object as distinct clusters at different azimuths (Fig. 5(d)). Finally, *Living Target Detection* isolates the static user's near-zero frequency band, enabling precise localization even in completely still scenes (Fig. 5(e)). Overall, this visualization highlights mmRehab's ability to recover fine-grained spatial awareness from residual clutter energy—transforming

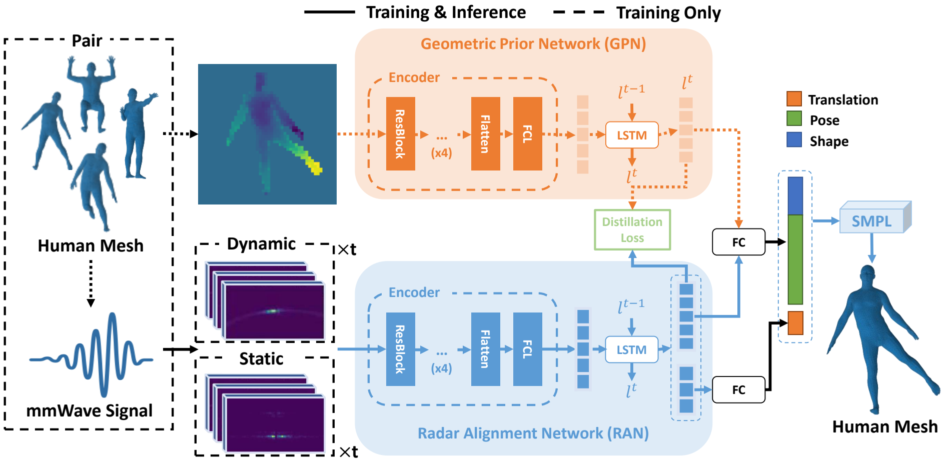


Fig. 6. The overview of our model.

weak echoes into interpretable static-user signatures that support downstream micro-motion characterization and environment mapping.

4.4 Geometry-aware Knowledge Transfer

Directly regressing human meshes from mmWave spectra is fundamentally ill-posed: radar reflections are sparse, noisy and lack the dense geometric structure that 3D meshes encode [49]. In contrast, depth maps provide spatially continuous geometry aligned with human body morphology, offering a stronger geometric prior. We therefore introduce **a geometry-aware knowledge distillation (KD) strategy** [28], using depth-based supervision [73] to guide the radar model toward learning generalizable geometric representations and mitigating device-specific overfitting [15, 79, 80].

As shown in Fig. 6, this module consists of two symmetric networks: a *Geometric Prior Network(GPN)* and a *Radar Alignment Network(RAN)*. Both adopt the same three-stage architecture: (1) a spatial encoder built with residual blocks [26] (denoted as ResBlocks) that extracts frame-level spatial features, (2) an LSTM [29] that captures temporal dynamics across frames, and (3) a fully connected regression head outputs the SMPL parameters. The GPN learns geometric priors from high-resolution depth images, while the RAN operates directly on mmWave spectra.

During training, cross-modal distillation transfers geometric knowledge from the GPN to the RAN by aligning their latent feature spaces. The RAN is optimized to reproduce the intermediate representations of the GPN via a mean-squared-error (MSE) loss, enabling it to internalize fine-grained spatial semantics despite the inherently low resolution of radar signals. This distillation process substantially improves robustness to signal noise and enhances generalization across unseen activities, users, and hardware configurations.

4.5 Human Mesh Reconstruction

A human mesh typically contain over 6,000 vertices, making direct prediction of all 3D vertex coordinates computationally intractable. To address this, we adopt the Skinned Multi-Person Linear (SMPL) [49] model, which represents a human body using only 82 parameters: 10 shape parameters Θ and 72 pose parameters β. Feeding these parameters into the SMPL decoder generates a mesh of 6890 vertices, allowing `mmRehab` to reconstruct detailed human geometry by regressing a compact parameter vector from mmWave features.

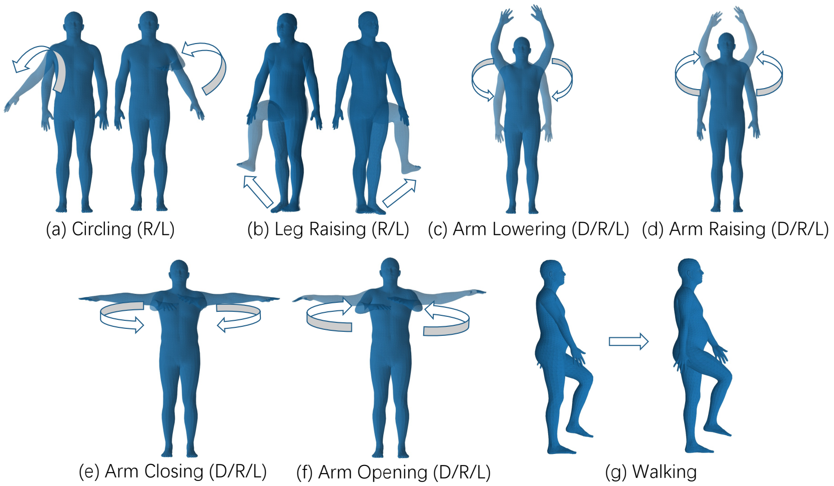


Fig. 7. Dynamic movements. This figure illustrates the set of dynamic rehabilitation actions captured by `mmRehab`, including arm and leg movements as well as walking. D/R/L denote performing the action with both, right, or left limbs, respectively.

To ensure anatomical continuity and numerical stability, we follow recent 3D reconstruction practice [92] and employ a continuous 6D rotation representation for joint orientations [92]. Specifically, the pose parameter θ is expanded to $\mathbb{R}^{144}$ (24 joints × 6D), and the resulting 6D vectors are orthogonalized into valid 3×3 matrices using a Gram-Schmidt process. This formulation eliminates discontinuities in rotation space and preserves differentiability during end-to-end training.

5 IMPLEMENTATION

This section provides the implementation details of the proposed `mmRehab` system, including datasets, loss functions, and model training configurations.

5.1 Datasets

To train and evaluate `mmRehab`, we combine measured and simulated mmWave data with synthetic motion sequences rendered from 3D meshes, achieving both realism and geometric consistency.

Synthetic Dataset. We utilize the AMASS dataset [50], a unified repository of human motion capture sequences expressed in SMPL parameters. For training, we use a diverse subset—including BMLrub, BMLhandball, CMU, DanceDB, GRAB, HUMAN4D, and KIT—covering 11,564 sequences from 310 subjects and approximately 2,328 minutes of motion. For evaluation, we select unseen sequences from DFaust, HumanEva, PosePrior, MoSh, SFU, SSM, TCDHands, TotalCapture, and Transitions (552 sequences from 52 subjects, 129 minutes). This broad coverage of poses, body shapes, and motion styles ensures strong generalization across unseen conditions and activities.

Real Dataset (Rehabilitation Scenarios). To capture realistic rehabilitation behaviors, we build a dedicated mmWave dataset containing 17 dynamic actions and 6 static postures, as illustrated in Fig. 7 and Fig. 8. Ten volunteers (6 males, 4 females; age 25.1 ± 2.43 years [range 22–29]; weight 69.3 ± 12.56 kg [range 45–83]; height 171.8 ± 9.11 cm [range 158–185]) participated, each performing every action for 70 seconds in randomized order to avoid sequence bias. This dataset bridges the gap between controlled synthetic motions and real-world sensing variations, covering both high-mobility and quasi-static rehabilitation conditions.

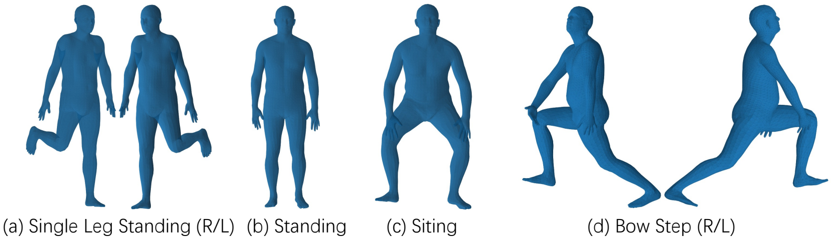


Fig. 8. Static postures. Representative static rehabilitation poses used for still-scene evaluation, including standing, sitting, bow-step, and single-leg support.

Table 1. The parameters of the mmWave radar system.

Parameter	**Value**
Start frequency	77 GHz
Sample points	128
Chirp loops	128
Maximum range	7.994 m
Range resolution	7.8 cm
Maximum velocity	3.767 m/s
Velocity resolution	5.9 cm/s
Frame periodicity	50 ms

Data were collected across three distinct environments—hall, meeting room, and living room—to capture environmental variability (Fig. 16). We use a TI IWR1843BOOST mmWave radar [32] paired with a TI DCA1000EVM adapter [31] for real-time signal streaming. The radar features 3 Tx and 4 Rx antennas; each Tx sequentially emits FMCW chirps while the Rx modules record reflections from the human subject and surrounding objects. Radar parameter settings are summarized in Table 1. The sensor is mounted at a height of 80 cm, colocated with an Orbbec Femto depth camera [53]) used for synchronized RGB capture. Groundtruth human meshes are generated using the VIBE model [38] applied to the camera recordings.

5.2 Losses

During the training of our model, we employ a multi-term loss functions to jointly supervise SMPL parameter regression, mesh consistency, temporal stability, and teacher-student feature alignment. The specific loss functions used are described below:

SMPL Parameters Loss. We predict SMPL parameters including translation ($\mathbf{p}_t \in \mathbb{R}^3$), pose ($\theta_t \in \mathbb{R}^{72}$), and body shape ($\beta_t \in \mathbb{R}^{10}$). The Mean Squared Error (MSE) is used as the loss function for these parameters:

$$\mathcal{L}_{\text{SMPL}} = \frac{1}{T}\sum_{t=1}^{T}\left(\|\mathbf{p}_t - \hat{\mathbf{p}_t}\|^2 + \|\mathbf{R}(\theta_t) - \mathbf{R}(\hat{\theta_t})\|^2 + \|\beta_t - \hat{\beta_t}\|^2\right)$$

where T is the sequnce length, $\mathbf{R}(\theta_i)$ represents the rotation matrix obtained from the pose parameters, and hats denote predictions.

Mesh Joints and Vertices Loss. The model computes the 3D coordinates of joints ($\mathbf{J}_t \in \mathbb{R}^{24\times3}$) and vertices ($\mathbf{V}_t \in \mathbb{R}^{6890\times3}$) derived from the SMPL parameters. The Mean Absolute Error (MAE) is utilized as the loss function for these coordinates:

$$\mathcal{L}_{\text{mesh}} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{n=1}^{N}\left(\|\mathbf{J}_t^n - \hat{\mathbf{J}}_t^n\|_1\right) + \frac{1}{M}\sum_{m=1}^{M}\left(\|\mathbf{V}_t^m - \hat{\mathbf{V}}_t^m\|_1\right)\right)$$

where N is the number of joints, M is the number of vertices, and $\|\cdot\|_1$ denotes the L1 norm.

Joints Velocity Loss. To avoid significant frame-to-frame prediction discrepancies which can result in jitter [34], the velocity of joints ($\mathbf{v}_{J,t} \in \mathbb{R}^{24\times3}$) is computed from consecutive frames. The predicted velocity is then subjected to an MSE loss:

$$\mathcal{L}_{\text{velocity}} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{v}_{J,t}^i - \hat{\mathbf{v}}_{J,t}^i\|^2$$

Distillation Loss. MSE is used to computes the error between the features from the teacher model ($\mathbf{l}_t$) and the student model ($\hat{\mathbf{l}_t}$):

$$\mathcal{L}_{\text{KD}} = \frac{1}{T}\sum_{t=1}^{T}\left(\|\mathbf{l}_t - \hat{\mathbf{l}}_t\|^2\right)$$

The final loss is a weighted summation of all the aforementioned losses:

$$\mathcal{L}_{\text{total}} = \alpha_s\mathcal{L}_{\text{SMPL}} + \alpha_m\mathcal{L}_{\text{mesh}} + \alpha_v\mathcal{L}_{\text{velocity}} + \alpha_k\mathcal{L}_{\text{KD}},$$

where α_s, α_m, α_v, and α_k are the weights for different losses.

5.3 Model Training

We implement all models using PyTorch [54] and train them on NVIDIA A800 GPUs.

The teacher model is trained on generated depth images with a resolution of 64 × 64. Its encoder comprises five ResBlocks, as described in [26], with filter sizes of 8, 16, 32, 64, and 128. The fully connected (FC) layer in the teacher model maps the flattened features into a 128-dimensional feature space, which is then transformed by an LSTM with 3 layers into a 192-dimensional feature space.

Similarly, the encoder of the student model consists of five ResBlocks with identical filter sizes (8, 16, 32, 64, and 128). Its FC layer also maps the flattened features to a 128-dimensional feature space, but the LSTM with 3 layers in the student model extends this to a 256-dimensional feature space. Of these, the first 192 dimensions are utilized for estimating pose and shape parameters, while the remaining 64 dimensions are utilized to estimating translation parameters.

We use a batch size of 16 to optimize memory usage and gradient stability, with the Adam optimizer chosen for its adaptive learning rate capabilities, initially set at 0.001. The weights for different losses, α_v, α_m, α_v, and α_k are empirically set to 0.2, 0.2, 1, and 0.2, respectively.

6 EVALUATION

To comprehensively evaluate the performance and practicality of `mmRehab`, we conduct extensive experiments under both controlled and real-world rehabilitation scenarios. This section introduces the experimental settings and evaluation metrics, followed by analyses of the system's overall performance, generalization robustness, and real-time inference capability.

6.1 Settings

We compare our system with two state-of-the-art methods for mmWave-based human mesh estimation and further conduct ablation studies to evaluate the contribution of each component under four settings:

- **M^{4}esh** [78]: Extends mesh reconstruction to multi-person scenarios by first detecting and tracking individual subjects via bounding boxes on a 2D MVDR-based energy map. For each detected subject, it generates a localized point cloud and reconstructs the mesh using a coarse-to-fine strategy. We adopt its single-person configuration for fair comparison.
- **mmGPE** [79]: A framework designed to generalize mesh estimation to unseen activities. It synthesizes realistic mmWave signals for novel poses and trains a mesh estimator jointly on real data (for seen activities) and synthetic data (for unseen ones).
- **Baseline (BSL):** Uses three radar spectra—range–azimuth, range–elevation, and range–Doppler—as input.
- **BSL + Static Clutters (BSL+SC):** Adds the static clutter spectrum to the baseline to capture stationary environmental reflections.
- **BSL + Static Feature Spectra (BSL+SFS):** Extends BSL by introducing three additional static-related spectra: static range–azimuth, static range–elevation, and fine-Doppler, enriching structural cues.
- **BSL + SFS + Knowledge Distillation (mmRehab):** Incorporates the geometry-aware knowledge distillation mechanism into BSL+SFS, forming the complete mmRehab framework.

6.2 Metrics

The following metrics are used to evaluate the performance of our system.

Average Vertex Error (V): This metric computes the mean Euclidean distance between corresponding vertices on the predicted and groundtruth meshes. It provides a direct measure of the geometric accuracy of the mesh model, reflecting how well the surface of the estimated mesh conforms to the actual human body captured in the dataset.

Average Joint Localization Error (S): This metric measures the average Euclidean distance between the predicted positions of joints and their actual positions in the groundtruth data. It quantifies the model's ability to accurately predict the location of human joints, which are pivotal for understanding body movements.

Procrustes Aligned Mean Per-Joint Position Error (PA-S): After aligning the predicted pose to the groundtruth pose using Procrustes analysis (which removes discrepancies caused by translation, rotation, and scaling), this metric evaluates the mean positional error per joint. It provides an insight into the intrinsic accuracy of the pose estimation, independent of global transformations.

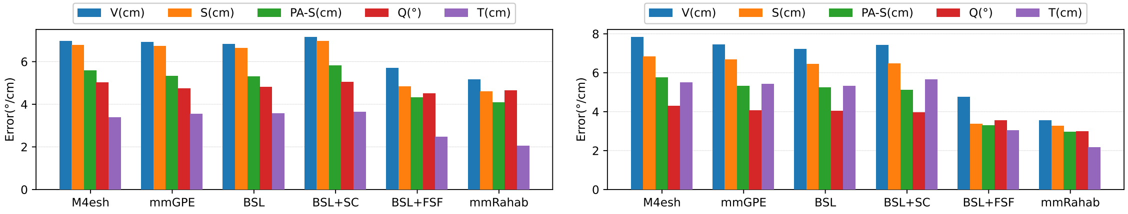
Average Joint Rotation Error (Q): This evaluates the average angular difference between the rotations of corresponding joints in the predicted and actual poses. It measures the model's capacity to capture the orientation of each joint correctly, which is crucial for dynamic activities that involve rotational movements.

Mesh Localization Error (T): Measures how accurately the entire estimated mesh is localized in space relative to the groundtruth mesh. This metric is calculated as the average distance between the centroids of the estimated and actual meshes and the spread of the error across the mesh surface.

6.3 Overall Performance

We first evaluate the overall performance of our system in two strong baselines and four different settings. Our model is trained on the first 60% of each sequence, validated on the following 20%, and tested on the last 20% from all subjects. The results are shown in Figure 9. Several conclusions can be drawn from the results:

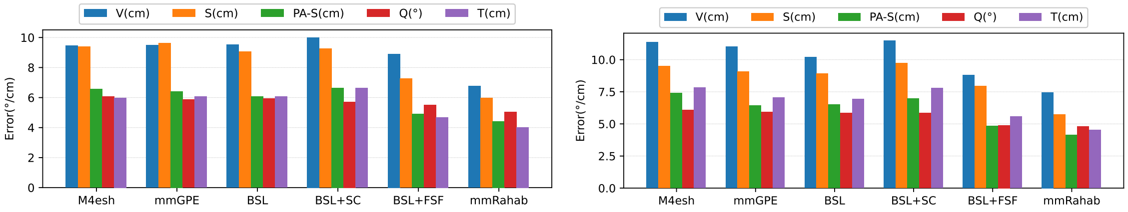
- **Comparison with Strong Baselines.** mmRehab consistently outperforms both M^{4}esh and mmGPE across all major metrics for both dynamic motions and static postures. For dynamic motions, mmRehab (+KD) reduces average vertex error (V) from 6.96 cm (M^{4}esh) to 5.16 cm, achieving a 25.9% improvement. Similar improvements are observed for average joint localization error (S, 32.0%), procrustes aligned mean per-Joint position error (PA-S, 23.2%), average joint rotation error (Q, 7.5%), and mesh localization error (T, 39.0%) compared to the best baseline. These results highlight that mmRehab achieves a superior balance between dynamic adaptability and



(a) Results for dynamic motions.

(b) Results for static postures.

Fig. 9. Overall performance comparison across dynamic (a) and static (b) scenarios. Bars indicate mean errors over five metrics—velocity (V), skeleton (S), pose-aligned skeleton (PA-S), joint angle (Q), and translation (T). mmRehab achieves the best performance under both conditions.



(a) Results for dynamic motions.

(b) Results for static postures.

Fig. 10. Generalization performance on unseen users. mmRehab consistently yields the lowest errors under both dynamic and quasi-static settings, demonstrating strong cross-user generalization in 3D mesh reconstruction.

static precision, with vitality-aware sensing and geometry-aware knowledge transfer enabling high-fidelity mesh reconstruction beyond existing mmWave-based frameworks.

- **Performance Across Different Settings.** Directly incorporating static clutter increases errors in both dynamic and static cases due to substantial background interference. In contrast, introducing fine-grained static feature spectra enables the model to better represent stationary body components, substantially improving mesh accuracy for both motion types. The improvement is especially evident for static postures, with errors reduced from (V: 7.21 cm, S: 6.44 cm, PA-S: 5.23 cm, Q: 4.03°, T: 5.32 cm) to (V: 4.75 cm, S: 3.38 cm, PA-S: 3.29 cm, Q: 3.54°, T: 3.03 cm). Further gains are achieved with knowledge distillation, which enhances latent feature representations and yields consistent improvements in both static and dynamic scenarios.

6.4 Generalization

To assess the generalization robustness of mmRehab, we further evaluate its performance on unseen users and unseen activities, ensuring that mmRehab can adapt to new subjects and motion patterns beyond those observed during training.

6.4.1 Unseen Users. In this section, we evaluate the generalization performance of our system on unseen users, which better reflects its effectiveness in real-world applications.

To simulate this scenario, we randomly selected three users (one female and two males) from a total of ten volunteers as unseen users, whose data were reserved for testing, while the remaining seven volunteers' data were used for model training.

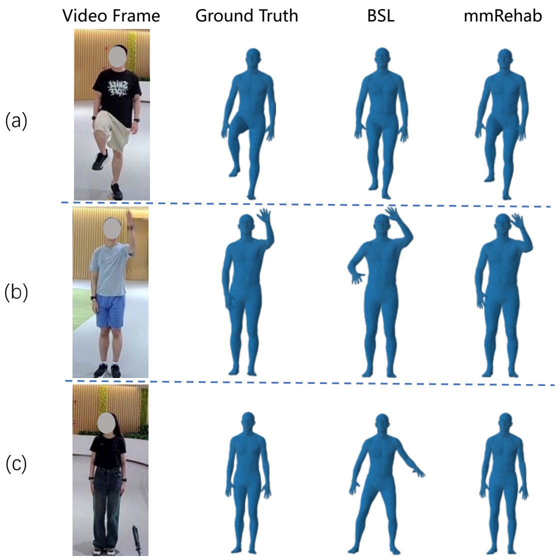


Fig. 11. Mesh estimation results for unseen users.

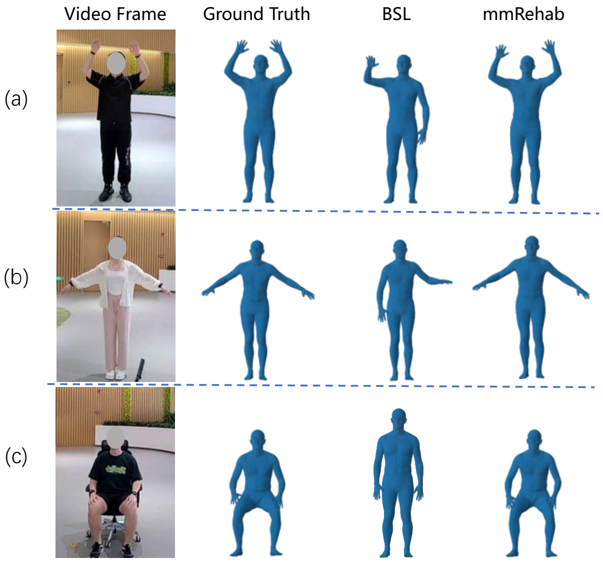


Fig. 12. Mesh estimation results for unseen postures.

The quantitative results are shown in Fig. 10. Although some performance degradation is inevitable on unseen users, our model still achieves accurate human mesh estimation for both dynamic motions (V: 6.77 cm, S: 5.97 cm, PA-S: 4.42 cm, Q: 5.04°, T: 4.03 cm) and static postures (V: 7.46 cm, S: 5.74 cm, PA-S: 4.13 cm, Q: 4.81°, T: 4.54 cm). This performance significantly outperforms strong baselines such as M^4esh and mmGPE.

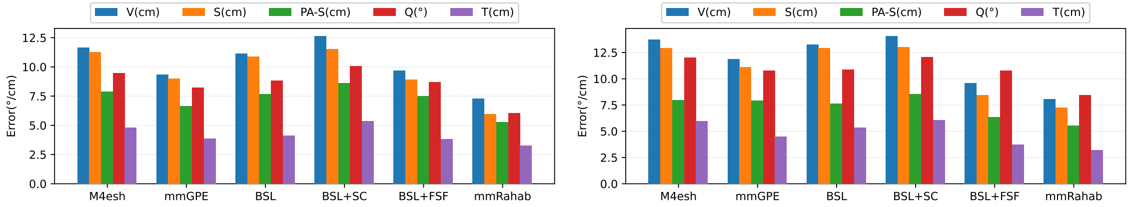
Examining the internal design variations highlights the contribution of each module. Compared to the baseline (BSL), directly incorporating static clutter (BSL+SC) does not consistently improve performance and even slightly increases errors due to environmental interference. Introducing fine-grained static feature spectra (BSL+SFS) substantially improves performance, reducing dynamic motion vertex error from 10.01 cm (BSL+SC) to 8.91 cm, and static posture vertex error from 11.49 cm (BSL+SC) to 8.80 cm. Finally, adding geometry-aware knowledge distillation (mmRahab) in mmRehab further reduces errors to 6.77 cm for dynamic motions and 7.46 cm for static postures, demonstrating that the combination of fine-grained static features and knowledge transfer is critical for robust generalization.

Figure 11 presents intuitive visual comparisons, including video frames, ground-truth (GT) human meshes, and estimated meshes under the BSL and mmRehab configurations.

- **Fig. 11 (a)**: The user performs a leg-lifting motion. While the BSL configuration detects the motion, it exhibits substantial errors in angle prediction. In contrast, mmRehab accurately predicts the leg-lifting angle.
- **Fig. 11 (b)**: The user lifts the left hand. Although BSL detects the left-hand motion, it produces significant errors for the stationary right hand. mmRehab, however, accurately reconstructs the poses of both hands.
- **Fig. 11 (c)**: The user maintains a static standing posture. The BSL configuration mispredicts the posture, whereas mmRehab again provides an accurate estimation.

These results indicate that, despite the inevitable challenges posed by unseen users, mmRehab maintains high-fidelity mesh reconstruction across all key metrics significantly outperforming both strong baselines and intermediate model variants, thus confirming superior generalization capability.

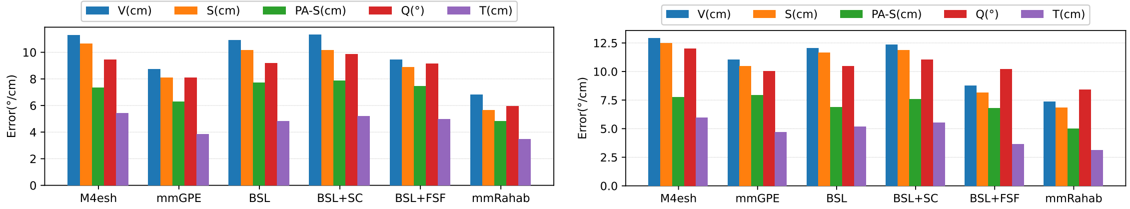
6.4.2 Unseen Activities. To evaluate the system's accuracy in recognizing unseen postures, we designed two distinct experimental setups, highlighting different levels of motion generalization.



(a) Results for Dynamic Motions.

(b) Results for Static Postures.

Fig. 13. Intra-limb unseen posture generalization.



(a) Results for Dynamic Motions.

(b) Results for Static Postures.

Fig. 14. Cross-limb unseen posture generalization.

- **Intra-Limb Unseen Postures.** In this setup, 7 out of 23 motions and postures from Figures 7 and 8 were selected as unseen for testing, while the remaining 16 were used for training. Specifically, dynamic cross-arm movements—*Raise Arms*, *Lower Arms*, *Open Arms*, and *Close Arms*—were treated as unseen postures. For static postures, *Bow Step* (right leg forward), *On One Leg* (right leg raised), and *Sit* were used as unseen examples. This setup evaluates generalization within similar motion categories (i.e., no cross-limb transfer).
- **Cross-Limb Unseen Postures.** To further assess generalization to fundamentally different motions, we tested cross-limb movements—*Open Arms* and *Walking*—that were not present in the training set. Static postures remained the same (*Bow Step*, *On One Leg*, and *Sit*). This setup introduces a larger domain gap and evaluates the model's ability to generalize across distinct limb categories.

Quantitative results are presented in Fig. 13 and 14 . The results indicate that mmGPE, which can generate realistic millimeter-wave signals for unseen postures, performs significantly better than all baselines except mmRehab on dynamic motions. However, its performance on critical static postures remains suboptimal. Directly incorporating SC further degrades system performance. The BSL+FSF configuration outperforms both BSL and BSL+SC, demonstrating that fine-grained static features provide more informative structural cues. By additionally applying knowledge distillation, mmRehab improves estimation accuracy for both unseen dynamic and static postures beyond BSL+FSF, highlighting its superior generalization capability.

For more intuitive analysis, Fig. 12 shows visual examples of a user raising both arms, closing both arms, and sitting, comparing the ground-truth and estimated meshes. While BSL struggles to capture the motion of both arms, mmRehab accurately predicts them. For the static sitting posture, mmRehab correctly reconstructs the pose, whereas BSL misclassifies it as standing. These examples clearly illustrate mmRehab's robustness in estimating unseen postures.

Table 2. Performance across unseen distances.

Distance	Dynamic Motions					Static Postures				
	V(cm)	S(cm)	PA-S(cm)	Q(°)	T(cm)	V(cm)	S(cm)	PA-S(cm)	Q(°)	T(cm)
2m	8.87	6.41	3.83	6.69	4.78	8.64	7.59	3.41	6.71	4.58
4m	7.46	6.79	5.23	6.36	2.35	10.47	9.40	6.47	9.30	4.15
5m	18.82	14.66	11.32	13.52	8.05	14.28	11.66	8.72	11.75	10.22
6m	18.64	14.22	11.05	13.75	8.04	17.80	12.47	8.54	15.77	11.81

Table 3. Performance across unseen orientations.

Angle	Dynamic Motions					Static Postures				
	V(cm)	S(cm)	PA-S(cm)	Q(°)	T(cm)	V(cm)	S(cm)	PA-S(cm)	Q(°)	T(cm)
30°	10.08	8.64	4.83	8.55	7.25	12.87	11.19	6.64	10.48	8.62
60°	12.75	11.78	6.78	10.30	10.46	16.53	15.13	10.47	11.91	11.70
90°	17.47	16.04	9.82	13.06	14.11	20.29	19.86	11.89	15.10	17.53

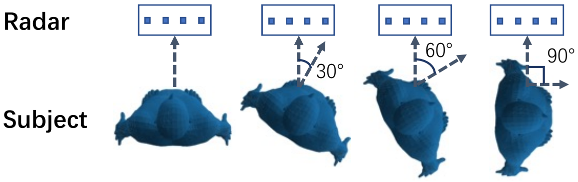


Fig. 15. Illustration of different radar–subject orientations.

6.5 Robustness

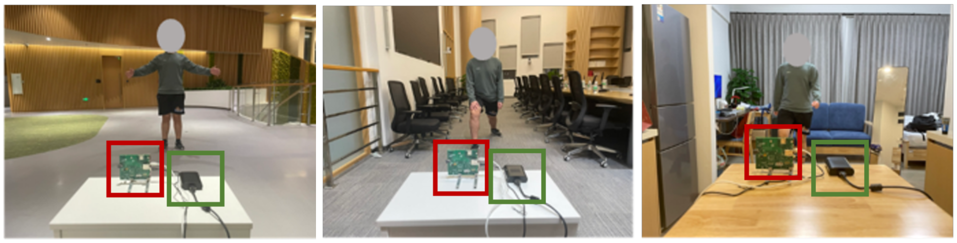
In this section, we evaluate the system's robustness from three aspects: distance, orientation, and environment. Specifically, for each distance, orientation, and environment, we collected data from four volunteers. Each volunteer is asked to perform each motion or posture for 70 seconds. The four volunteers are unchanged throughout the experiment. We use the model trained in the overall performance evaluation to test on these newly collected data, thereby assessing the system's performance on unseen distances, orientations, and environments.

6.5.1 Impact of Unseen Distances. We collect data from four volunteers at four additional distances: 2m, 4m, 5m, and 6m. The Table 2 shows the results for different distances. We can see that when the distance is 2m and 4m, the system can still accurately estimate the human mesh, demonstrating good robustness across various distances. However, when the distance increases to 5m and beyond, the system's performance declines dramatically. The degradation can be attributed to two primary factors. Firstly, the reflected signal from the human body weakens at greater distances, leading to a lower signal-to-noise ratio (SNR). Secondly, the human subject occupies a smaller region in the radar heatmap, resulting in a loss of spatial detail and reduced feature separability. The impact of these effects on the input distribution at longer distances is such that it becomes significantly different from the training data collected at near range. This, in turn, has a detrimental effect on reconstruction accuracy.

6.5.2 Impact of Unseen Orientations. We collected data with volunteers facing the radar at different angles. While the standard setup uses 0°, real-world scenarios often involve users not directly facing the radar. To evaluate system performance under such conditions, we additionally collected data at 30°, 60°, and 90° (Figure 15). Quantitative results are presented in Table 3.

Table 4. Performance across unseen environments.

Environment	Dynamic Motions					Static Postures				
	V(cm)	S(cm)	PA-S(cm)	Q(°)	T(cm)	V(cm)	S(cm)	PA-S(cm)	Q(°)	T(cm)
Meeting Room	8.82	7.67	5.70	7.03	5.65	10.17	9.22	6.65	8.33	6.49
Living Room	10.38	9.79	6.82	9.17	6.67	10.84	9.68	7.59	9.52	6.82



(a) Hall (b) Meeting Room (c) Living Room

Fig. 16. Experimental setups across different environments. (a) Hall, (b) Meeting Room, and (c) Living Room represent distinct indoor scenes with varying background clutter, furniture layout, and multipath conditions. The radar and the camera are highlighted by red and green boxes, respectively.

At 30°, the system still accurately estimates the human mesh. When the angle increases to 60°, performance shows a noticeable decline, and at 90°, accuracy drops dramatically. This degradation is primarily due to two factors: (1) changes in the reflection pattern caused by varying angles, which affect signal quality, and (2) a reduction in the effective reflective surface area of the human body as the angle increases, limiting the information available for accurate mesh estimation.

In future work, this issue could be mitigated by deploying radars at multiple positions and orientations to capture the human body from diverse viewpoints, thereby improving robustness to user orientation.

6.5.3 Impact of Unseen Environments. To evaluate the system's performance in previously unseen environments, we collected data from four volunteers in two additional settings: a meeting room and a living room (Figure 16). These environments are more complex, containing multiple surrounding objects that pose additional challenges for human mesh estimation. Quantitative results are summarized in Table 4.

Despite the increased environmental complexity, the system achieves accurate human mesh reconstruction in both settings, demonstrating strong robustness and generalizability to diverse, previously unseen environments.

6.6 Real-time Inference Latency

To evaluate the real-time performance, we analyze the computational latency of the mmRehab system. We conduct the inference on a desktop (Intel(R) Core(TM) i7-14700KF CPU, NVIDIA GeForce RTX 4090 D GPU) and report the average latency. The whole inference process is divided into five main stages: (1) Data Collection, which includes capturing the raw IF signal from the radar and transferring it to the host machine; (2) Hierarchical Feature Extraction, where the raw signal is processed by both the Dynamic Spatial-Kinematic Mapping and Static Vital Micro-motion Extraction modules to generate the spatial-physiological representations; (3) Mesh Estimation, which involves feeding the extracted features into the Radar Alignment Network to predict the SMPL parameters; (4) Mesh Generation, which decodes the predicted parameters via the SMPL model to generate the 3D mesh vertices; and (5) Visualization of the final reconstructed human mesh.

Their average latencies are 105.2 ms, 174.7 ms, 9.5 ms, 3.3 ms, and 38.6 ms, respectively. In total, the end-to-end inference latency of the whole system is 331.3 ms, which demonstrates that our proposed system can operate in real-time for continuous rehabilitation monitoring.

7 DISCUSSION

This section first discusses the remaining challenges and future directions of `mmRehab`, followed by considerations on ethics and privacy.

7.1 Limitations and Future Work

While `mmRehab` demonstrates robust performance across diverse postures and users, two limitations remains.

- **Angular Sensitivity.** The system's accuracy decreases under extreme incidence angles, especially near 90 degrees, where only a limited portion of the body surface is visible to the radar, leading to insufficient geometric cues for reconstruction. This limitation can be mitigated by deploying multiple radars to achieve full-body coverage and by expanding the antenna array to enhance angular resolution and reduce view dependence.
- **Single-user Constraint.** Since `mmRehab` lacks segmentation and identity tracking, it currently supports only single-user sensing. Extending it to multi-user environments—where echoes from multiple bodies interfere—is an important next step, motivated by recent advances in multi-person mmWave tracking [14, 76, 78].

7.2 Ethics and Privacy

All experiments involving human participants were reviewed and approved by the institutional review board (IRB). Prior to data collection, participants in both the main experiments and downstream evaluations were informed of the study's objectives, procedures, and potential risks, and each provided written informed consent. During data processing, all personally identifiable information (PII) was anonymized or removed. Physiological and behavioral data were stored on encrypted servers with restricted access and used solely for research purposes. The data will not be publicly released without additional anonymization and renewed participant consent.

8 CONCLUSION

In this work, we addressed a fundamental limitation of RF sensing—its reliance on motion cues, which makes it blind to the stillness that characterizes balance, endurance, and postural control in rehabilitation. We introduced `mmRehab`, a unified mmWave sensing system that bridges motion-centric sensing and life-state perception. The key novelty of `mmRehab` is to re-interpret what radar once dismissed as "clutter" as a rich source of information: the reflections of micro-motions—minute signatures of life such as breathing, heartbeat, and subtle sway—encode physiological vitality even in apparent stillness. `mmRehab` integrates two synergistic components: i) the *Micro-motion Feature Extraction* couples beamforming-based spatial isolation with micro-Doppler temporal discrimination to preserve and amplify respiration- and posture-related cues that are otherwise buried within near-zero Doppler clutter; and ii) the *Geometry-aware Knowledge Transfer* leverages depth maps as an intemediate bridge to transfer structural priors from vision to radar, enabling fine-grained 3D human mesh reconstruction. Extensive experiments across dynamic and static rehabilitation tasks show that `mmRehab` reduces 3D reconstruction errors by over 24% and generalizes robustly to unseen users, distances, and orientations. Overall, `mmRehab` transforms mmWave radar from a motion detector into a **life-state observer**, marking a step toward continuous, contactless, and clinically meaningful rehabilitation monitoring.

ACKNOWLEDGMENTS

This research is partially supported by China NSFC Grant (62472366), the Project of DEGP (No.2023KCXTD042, 2024GCZX003), Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No.2023B1212010007), "111 Center (No.D25008)", Shenzhen Science and Technology Foundation (ZDSYS20190902092853047).

References

[1] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–13.

[2] Radzi Bin Ambar, Hazwaj Bin Mhd Poad, Abdul Malik Bin Mohd Ali, Muhammad Shukri Bin Ahmad, and Muhammad Mahadi Bin Abdul Jamil. 2012. Multi-sensor arm rehabilitation monitoring device. In *2012 International Conference on Biomedical Engineering (ICoBE)*. IEEE, 424–429.

[3] Sizhe An, Yin Li, and Umit Ogras. 2022. mri: Multi-modal 3d human pose estimation dataset using mmwave, rgb-d, and inertial sensors. *Advances in neural information processing systems* 35 (2022), 27414–27426.

[4] Sizhe An and Umit Y Ogras. 2021. Mars: mmwave-based assistive rehabilitation system for smart healthcare. *ACM Transactions on Embedded Computing Systems (TECS)* 20, 5s (2021), 1–22.

[5] Donald E Barrick. 1973. *FM/CW radar signals and digital processing*. Vol. 55. Environmental Research Laboratories.

[6] Matteo Bastico, Alberto Belmonte-Hernández, and Federico Álvarez García. 2022. Continuous Person Identification and Tracking in Healthcare by Integrating Accelerometer Data and Deep Learning Filled 3D Skeletons. *IEEE Sensors Journal* 22, 15 (2022), 15402–15409.

[7] Rosa Cabanas-Valdes, Gerard Urrutia Cuchi, and Caritat Bagur-Calafat. 2013. Trunk training exercises approaches for improving trunk performance and functional sitting balance in patients with stroke: a systematic review. *NeuroRehabilitation* 33, 4 (2013), 575–592.

[8] Zhongping Cao, Guangyu Mei, Xuemei Guo, and Guoli Wang. 2024. VirTeach: mmWave Radar Point Cloud Based Pose Estimation With Virtual Data as a Teacher. *IEEE Internet of Things Journal* (2024).

[9] Jack Capon. 1969. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* 57, 8 (1969), 1408–1418.

[10] Zhaoxin Chang, Fusang Zhang, Xujun Ma, Pei Wang, Weiyan Chen, Duo Zhang, Badii Jouaber, and Daqing Zhang. 2024. MmECare: Enabling Fine-grained Vital Sign Monitoring for Emergency Care with Handheld MmWave Radars. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 207 (Nov. 2024), 24 pages. doi:10.1145/3699766

[11] Anjun Chen, Xiangyu Wang, Kun Shi, Yuchi Huo, Jiming Chen, and Qi Ye. 2025. Toward Weather-Robust 3D Human Body Reconstruction: Millimeter-Wave Radar-Based Dataset, Benchmark, and Multi-Modal Fusion. *IEEE Transactions on Circuits and Systems for Video Technology* 35, 1 (2025), 273–286. doi:10.1109/TCSVT.2024.3461960

[12] Anjun Chen, Xiangyu Wang, Kun Shi, Shaohao Zhu, Bin Fang, Yingfeng Chen, Jiming Chen, Yuchi Huo, and Qi Ye. 2023. Immfusion: Robust mmwave-rgb fusion for 3d human body reconstruction in all weather conditions. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2752–2758.

[13] Wenqiang Chen, Yexin Hu, Wei Song, Yingcheng Liu, Antonio Torralba, and Wojciech Matusik. 2024. CAvatar: Real-time Human Activity Mesh Reconstruction via Tactile Carpets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–24.

[14] Weiyan Chen, Hongliu Yang, Xiaoyang Bi, Rong Zheng, Fusang Zhang, Peng Bao, Zhaoxin Chang, Xujun Ma, and Daqing Zhang. 2023. Environment-aware Multi-person Tracking in Indoor Environments with MmWave Radars. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 89 (Sept. 2023), 29 pages. doi:10.1145/3610902

[15] Xingyu Chen and Xinyu Zhang. 2023. RF Genesis: Zero-shot generalization of mmwave sensing through simulation-based data synthesis and generative diffusion models. In *ACM Conference on Embedded Networked Sensor Systems (SenSys' 23)*.

[16] Henry M Clever, Patrick L Grady, Greg Turk, and Charles C Kemp. 2022. Bodypressure-inferring body pose and contact pressure from a depth image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 1 (2022), 137–153.

[17] Saverio Colonna, Antonio D'Alessandro, Riccardo Tarozzi, and Fabio Casacci. 2025. Supine bridge exercise: a narrative review of the literature (Part I). *Cureus* 17, 3 (2025).

[18] Anne De Groote, Muriel Wantier, Guy Chéron, Marc Estenne, and Manuel Paiva. 1997. Chest wall motion during tidal breathing. *Journal of Applied Physiology* 83, 5 (1997), 1531–1537.

[19] Han Ding, Zhenbin Chen, Cui Zhao, Fei Wang, Ge Wang, Wei Xi, and Jizhong Zhao. 2023. MI-Mesh: 3D human mesh construction by fusing image and millimeter wave. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 1 (2023), 1–24.

[20] Deborah Edwards, Jenny Williams, Judith Carrier, and Jennifer Davies. 2022. Technologies used to facilitate remote rehabilitation of adults with deconditioning, musculoskeletal conditions, stroke, or traumatic brain injury: an umbrella review. *JBI evidence synthesis* 20, 8 (2022), 1927–1968.

[21] Min Yong Eom, Sin Ho Chung, and Tae Sung Ko. 2013. Effects of bridging exercise on different support surfaces on the transverse abdominis. *Journal of physical therapy science* 25, 10 (2013), 1343–1346.

[22] Pragati Garg, Naveen Aggarwal, and Sanjeev Sofat. 2009. Vision based hand gesture recognition. *International Journal of Computer and Information Engineering* 3, 1 (2009), 186–191.

[23] Jiaqi Geng, Dong Huang, and Fernando De la Torre. 2022. Densepose from wifi. *arXiv preprint arXiv:2301.00250* (2022).

[24] Moran Gilat, Pieter Ginis, Demi Zoetewei, Joni De Vleeschhauwer, Femke Hulzinga, Nicholas D'Cruz, and Alice Nieuwboer. 2021. A systematic review on exercise and training-based interventions for freezing of gait in Parkinson's disease. *NPJ Parkinson's disease* 7, 1 (2021), 81.

[25] Shubham Goel, Georgios Pavlakos, Jathushan Rajasegaran, Angjoo Kanazawa, and Jitendra Malik. 2023. Humans in 4D: Reconstructing and tracking humans with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14783–14794.

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[27] Steven M Hernandez, Md Touhiduzzaman, Peter E Pidcoe, and Eyuphan Bulut. 2022. Wi-PT: Wireless sensing based low-cost physical rehabilitation tracking. In *2022 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*. IEEE, 113–118.

[28] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the Knowledge in a Neural Network. arXiv:1503.02531 [stat.ML] https://arxiv.org/abs/1503.02531

[29] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[30] Yo-Ping Huang, Yu-Yu Liu, Wei-Hsiu Hsu, Li-Ju Lai, and Mel S Lee. 2020. Monitoring and assessment of rehabilitation progress on range of motion after total knee replacement by sensor-based system. *Sensors* 20, 6 (2020), 1703.

[31] Texas Instruments. 2024. DCA1000EVM. https://www.ti.com/tool/DCA1000EVM.

[32] Texas Instruments. 2024. IWR1843BOOST. https://www.ti.com/product/IWR1843BOOST/part-details/IWR1843BOOST.

[33] Texas Instruments. 2024. mmWave Sensing Estimator. https://dev.ti.com/gallery/view/mmwave/mmWaveSensingEstimator/ver/2.4.0/.

[34] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun, and Lu Su. 2020. Towards 3D human pose construction using WiFi. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

[35] Charles H Jones and Mikael Dolsten. 2024. Healthcare on the brink: navigating the challenges of an aging society in the United States. *npj Aging* 10, 1 (2024), 22.

[36] Angjoo Kanazawa, Michael J Black, David W Jacobs, and Jitendra Malik. 2018. End-to-end recovery of human shape and pose. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7122–7131.

[37] Sagi Katz, Ayellet Tal, and Ronen Basri. 2007. Direct visibility of point sets. In *ACM SIGGRAPH 2007 Papers* (San Diego, California) *(SIGGRAPH '07)*. Association for Computing Machinery, New York, NY, USA, 24–es. doi:10.1145/1275808.1276407

[38] Muhammed Kocabas, Nikos Athanasiou, and Michael J Black. 2020. Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5253–5263.

[39] Nikos Kolotouros, Georgios Pavlakos, Michael J Black, and Kostas Daniilidis. 2019. Learning to reconstruct 3D human pose and shape via model-fitting in the loop. In *Proceedings of the IEEE/CVF international conference on computer vision*. 2252–2261.

[40] Hao Kong, Xiangyu Xu, Jiadi Yu, Qilin Chen, Chenguang Ma, Yingying Chen, Yi-Chao Chen, and Linghe Kong. 2022. m3track: mmwave-based multi-user 3d posture tracking. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*. 491–503.

[41] Jan Kottner, Janet Cuddigan, Keryln Carville, Katrin Balzer, Dan Berlowitz, Susan Law, Mary Litchford, Pamela Mitchell, Zena Moore, Joyce Pittman, et al. 2019. Prevention and treatment of pressure ulcers/injuries: The protocol for the second update of the international Clinical Practice Guideline 2019. *Journal of tissue viability* 28, 2 (2019), 51–58.

[42] Wojciech Kurzydło, Beata Stach, Aleksandra Bober, Mariola E Wodzińska, and Mirosława M Długosz. 2014. Structured-light 3D scanner in use to access the human body posture in physical therapy: a pilot study. (2014).

[43] Jiale Lai, Jiake Tian, Yi Zou, Xianfeng Song, Fangming Liu, and Dacheng Li. 2024. mmHPE: Human Pose Estimation Based on Point Cloud from Millimeter-wave Radar. In *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2325–2331.

[44] Eun-Lee Lee, Myung Hun Jang, Byeong-Ju Lee, Sang Hun Han, Hoo Man Lee, Sang Ui Choi, and Myung Jun Shin. 2023. Home-based remote rehabilitation leads to superior outcomes for older women with knee osteoarthritis: a randomized controlled trial. *Journal of the American Medical Directors Association* 24, 10 (2023), 1555–1561.

[45] Gun-Hee Lee and Seong-Whan Lee. 2021. Uncertainty-aware human mesh recovery from video by learning part-based 3d dynamics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12375–12384.

[46] Jian Li and Petre Stoica. 2005. *Robust adaptive beamforming*. John Wiley & Sons.

[47] Saiyi Li, Pubudu N Pathirana, and Terry Caelli. 2014. Multi-kinect skeleton fusion for physical rehabilitation monitoring. In *2014 36th Annual international conference of the IEEE engineering in medicine and biology society*. IEEE, 5060–5063.

[48] Wilburta Q Lindh, Marilyn Pooler, Carol Tamparo, Barbara M Dahl, and J Morris. 2009. *Delmar's Comprehensive Medical Assisting: Ad-ministrative and Clinical Competencies. Cengage Learning.* Technical Report. ISBN 978-1-4354-1914-8.

[49] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.

[50] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. 2019. AMASS: Archive of Motion Capture as Surface Shapes. In *International Conference on Computer Vision*. 5442–5451.

[51] José-María Muñoz-Ferreras, Guochao Wang, Changzhi Li, and Roberto Gómez-García. 2015. Mitigation of stationary clutter in vital-sign-monitoring linear-frequency-modulated continuous-wave radars. *IET Radar, Sonar & Navigation* 9, 2 (2015), 138–144.

[52] Bennett K Ng, Benjamin J Hinton, Bo Fan, Alka M Kanaya, and John A Shepherd. 2016. Clinical anthropometrics and body composition from 3D whole-body surface scans. *European journal of clinical nutrition* 70, 11 (2016), 1265–1270.

[53] Orbbec. 2024. Femto. https://www.orbbec.com/products/archived-products/femto/.

[54] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).

[55] Octavian Postolache, D Jude Hemanth, Ricardo Alexandre, Deepak Gupta, Oana Geman, and Ashish Khanna. 2020. Remote monitoring of physical rehabilitation of stroke patients using IoT and virtual reality. *IEEE Journal on Selected Areas in Communications* 39, 2 (2020), 562–573.

[56] Jathushan Rajasegaran, Georgios Pavlakos, Angjoo Kanazawa, and Jitendra Malik. 2022. Tracking people by predicting 3d appearance, location and pose. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2740–2749.

[57] Sandeep Rao. 2017. Introduction to mmWave sensing: FMCW radars. *Texas Instruments (TI) mmWave Training Series* (2017), 1–11.

[58] Yili Ren, Zi Wang, Sheng Tan, Yingying Chen, and Jie Yang. 2021. Winect: 3d human pose tracking for free-form activity using commodity wifi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–29.

[59] Emilio Sardini, Mauro Serpelloni, and Viviane Pasqui. 2014. Wireless wearable T-shirt for posture monitoring during rehabilitation exercises. *IEEE Transactions on Instrumentation and Measurement* 64, 2 (2014), 439–448.

[60] Louis L Scharf. 2004. Statistical signal processing: applications to beamforming, detection, and estimation. In *Medical Imaging 2004: Ultrasonic Imaging and Signal Processing*, Vol. 5373. SPIE, 1–17.

[61] Tanja Schwarz, Andrea E Schmidt, Julia Bobek, and Joy Ladurner. 2022. Barriers to accessing health care for people with chronic conditions: a qualitative interview study. *BMC health services research* 22, 1 (2022), 1037.

[62] Arindam Sengupta, Feng Jin, Renyuan Zhang, and Siyang Cao. 2020. mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs. *IEEE Sensors Journal* 20, 17 (2020), 10032–10044.

[63] Cong Shi, Li Lu, Jian Liu, Yan Wang, Yingying Chen, and Jiadi Yu. 2022. mPose: Environment-and subject-agnostic 3D skeleton posture reconstruction leveraging a single mmWave device. *Smart Health* 23 (2022), 100228.

[64] Nicola Smania, Elisabetta Corato, Michele Tinazzi, Clementina Stanzani, Antonio Fiaschi, Paolo Girardi, and Marialuisa Gandolfi. 2010. Effect of balance training on postural instability in patients with idiopathic Parkinson's disease. *Neurorehabilitation and neural repair* 24, 9 (2010), 826–834.

[65] Petre Stoica, Randolph L Moses, et al. 2005. *Spectral analysis of signals*. Vol. 452. Pearson Prentice Hall Upper Saddle River, NJ.

[66] Scott Telfer and James Woodburn. 2010. The use of 3D surface scanning for the measurement and assessment of the human foot. *Journal of foot and ankle research* 3, 1 (2010), 19.

[67] Jiake Tian, Yi Zou, and Jiale Lai. 2025. mmHSE: A Two-Stage Framework for Human Skeleton Estimation Using mmWave FMCW Radar Signals. *Applied Sciences* 15, 15 (2025). doi:10.3390/app15158410

[68] Md Touhiduzzaman, Steven M. Hernandez, Peter E. Pidcoe, and Eyuphan Bulut. 2025. Wi-PT-Hand: Wireless Sensing based Low-cost Physical Rehabilitation Tracking for Hand Movements. *ACM Trans. Comput. Healthcare* 6, 1, Article 12 (Jan. 2025), 25 pages. doi:10.1145/3688855

[69] Harry L Van Trees. 2002. *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons.

[70] Logan Wade, Laurie Needham, Polly McGuigan, and James Bilzon. 2022. Applications and limitations of current markerless motion capture methods for clinical gait biomechanics. *PeerJ* 10 (2022), e12995.

[71] Fei Wang, Stanislav Panev, Ziyi Dai, Jinsong Han, and Dong Huang. 2019. Can WiFi estimate person pose? *arXiv preprint arXiv:1904.00277* (2019).

[72] Fei Wang, Sanping Zhou, Stanislav Panev, Jinsong Han, and Dong Huang. 2019. Person-in-WiFi: Fine-grained person perception using WiFi. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5452–5461.

[73] Yan Wang, Wei-Lun Chao, Divyansh Garg, Bharath Hariharan, Mark Campbell, and Kilian Q Weinberger. 2019. Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8445–8453.

[74] WHO. 2024. Rehabilitation. https://www.who.int/news-room/fact-sheets/detail/rehabilitation.

[75] Jonathan Williams, Jimmy Wang, Michael Stevens, Nigel Lovell, and Ahmadreza Argha. 2024. mmYOLOH-p: A Clinically-Oriented mmWave-Based Human Pose Estimation Tool for Unobtrusive Patient Monitoring. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 1–4.

[76] Chenshu Wu, Feng Zhang, Beibei Wang, and K. J. Ray Liu. 2020. mmTrack: Passive Multi-Person Localization Using Commodity Millimeter Wave Radio. In *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*. 2400–2409. doi:10.1109/INFOCOM41043.2020.9155293

[77] Xiangyu Xu, Lijuan Liu, and Shuicheng Yan. 2023. SMPLer: Taming Transformers for Monocular 3D Human Shape and Pose Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[78] Hongfei Xue, Qiming Cao, Yan Ju, Haochen Hu, Haoyu Wang, Aidong Zhang, and Lu Su. 2022. M4esh: mmwave-based 3d human mesh construction for multiple subjects. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 391–406.

[79] Hongfei Xue, Qiming Cao, Chenglin Miao, Yan Ju, Haochen Hu, Aidong Zhang, and Lu Su. 2023. Towards generalized mmwave-based human pose estimation through signal augmentation. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–15.

[80] Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. 2021. mmMesh: Towards 3D real-time dynamic human mesh construction using millimeter-wave. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. 269–282.

[81] Zhicheng Yang, Parth H Pathak, Yunze Zeng, Xixi Liran, and Prasant Mohapatra. 2016. Monitoring vital signs using millimeter wave. In *Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing*. 211–220.

[82] Rasoul Yousefi, Sarah Ostadabbas, Miad Faezipour, Masoud Farshbaf, Mehrdad Nourani, Lakshman Tamil, and Matthew Pompeo. 2011. Bed posture classification for pressure ulcer prevention. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 7175–7178.

[83] Chenglong Yu, Xuan Wang, Hejiao Huang, Jianping Shen, and Kun Wu. 2010. Vision-based hand gesture recognition using combinational features. In *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. IEEE, 543–546.

[84] Bo Zhang, Boyu Jiang, Rong Zheng, Xiaoping Zhang, Jun Li, and Qiang Xu. 2023. Pi-vimo: Physiology-inspired robust vital sign monitoring using mmwave radars. *ACM Transactions on Internet of Things* 4, 2 (2023), 1–27.

[85] Jia Zhang, Rui Xi, Yuan He, Yimiao Sun, Xiuzhen Guo, Weiguo Wang, Xin Na, Yunhao Liu, Zhenguo Shi, and Tao Gu. 2023. A survey of mmWave-based human sensing: Technology, platforms and applications. *IEEE Communications Surveys & Tutorials* 25, 4 (2023), 2052–2087.

[86] Xiaotong Zhang, Zhenjiang Li, and Jin Zhang. 2022. Synthesized millimeter-waves for human motion sensing. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 377–390.

[87] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-wall human pose estimation using radio signals. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7356–7365.

[88] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. 2018. RF-based 3D skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. 267–281.

[89] Wenbing Zhao, Hai Feng, Roanna Lun, Deborah D Espy, and M Ann Reinthal. 2014. A Kinect-based rehabilitation exercise monitoring and guidance system. In *2014 IEEE 5th international conference on software engineering and service science*. IEEE, 762–765.

[90] Ce Zheng, Xianpeng Liu, Mengyuan Liu, Tianfu Wu, Guo-Jun Qi, and Chen Chen. 2023. DiffMesh: A Motion-aware Diffusion-like Framework for Human Mesh Recovery from Videos. *arXiv preprint arXiv:2303.13397* (2023).

[91] Xiaolong Zhou, Tian Jin, Yongpeng Dai, Yongping Song, and Kemeng Li. 2024. Three-Dimensional Human Pose Estimation from Micro-Doppler Signature Based on SISO UWB Radar. *Remote Sensing* 16, 7 (2024). doi:10.3390/rs16071295

[92] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. 2019. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5745–5753.