Expanding mmWave Datasets for Human Pose Estimation with Unlabeled Data and LiDAR Datasets

Zhuoxuan Peng Boan Zhu Xingjian Zhang Wenying Li S.-H. Gary Chan
The Hong Kong University of Science and Technology
zpengac@cse.ust.hk, {bzhual, xzhangha, wlidt}@connect.ust.hk, gchan@cse.ust.hk

Abstract

Current millimeter-wave (mmWave) datasets for human pose estimation (HPE) are scarce and lack diversity in both point cloud (PC) attributes and human poses, hindering the generalization ability of their trained models. On the other hand, unlabeled mmWave HPE data and diverse LiDAR HPE datasets are readily available. We propose EMDUL, a novel approach to **e***xpand the volume and diversity of an existing* **m***mWave* **d***ataset using* **u***nlabeled mmWave data and* **L***iDAR datasets. EMDUL consists of two independent modules, namely a pseudo-label estimator to annotate unlabeled mmWave data, and a closed-form converter that translates an annotated LiDAR PC to its mmWave counterpart. Expanding the original dataset with both LiDAR-converted and pseudo-labeled mmWave PCs, schname significantly boosts the performance and generalization ability of all the examined HPE models, reducing 15.1% and 18.9% error for in-domain and out-of-domain settings, respectively. Code is available at* https://github.com/Shimmer93/EMDUL.

1. Introduction

Human pose estimation (HPE) is to predict the human skeleton in terms of the locations and connectivity of the joints (or keypoints). It has broad applications in robotics, human-computer interaction, action recognition, etc. Millimeter-wave (mmWave) HPE has drawn wide and sustained attention in recent years due to the strengths over traditional RGB cameras in terms of its 3D nature, user privacy protection, robustness against lighting conditions, etc. For training and inference purposes, mainstream mmWave HPE has adopted point cloud (PC) due to its representation simplicity and processing efficiency.

Despite the long-standing community interest in mmWave HPE, at present its labeled sets of data, the so-called "datasets," are still scarce. Furthermore, their diversity is limited in two critical aspects: (1) *PC attributes*, such as detection noise, point density, and human motion sensitivity, mainly due to lack of device heterogeneity and environmental variety during data collection; and (2) *Human poses*, where subjects often move in simple postures facing the radar with marked uniformity. Such datasets severely undermine model generalization, leading to degraded performance in real-world or unseen environments.

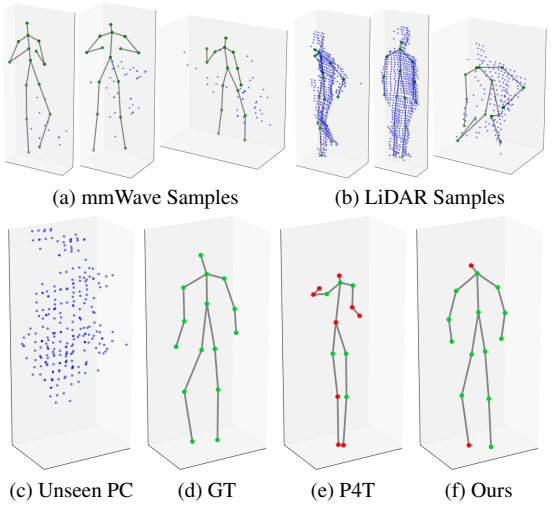


Figure 1. Examples illustrating the effect of dataset expansion. (a) Samples from an mmWave HPE training dataset. (b) Samples from a LiDAR dataset with richer pose diversity used for dataset expansion; (c) An mmWave PC from an unseen scenario. (d) The ground-truth skeleton. (e) The predicted skeleton of SOTA P4T [11] without expansion. (f) The predicted skeleton of P4T trained on EMDUL-expanded dataset. Joints are colored red for errors $>$ 10 cm and green otherwise. EMDUL achieves stronger generalization ability than the baseline P4T.

While labeled mmWave data is scarce, *unlabeled* mmWave data with diverse poses can be easily collected with minimal setup and manual annotations. The key issue is how to annotate automatically, i.e., "pseudo-label", such unlabeled data to embrace them into the current mmWave datasets to enhance the data diversity for training.

We further note that the PC datasets of LiDAR, another prevalent sensor for HPE, are abundant and widely available, often covering more diverse poses than mmWave datasets. It would be desirable to leverage these LiDAR datasets for mmWave training. However, this goal is challenging to achieve because the intrinsic PC attributes of LiDAR are fundamentally different from those of mmWave (due to distinct physical principles of sensing). We need to bridge the gap between the two modalities to make the LiDAR datasets useful for mmWave training.

To expand the existing mmWave datasets, we investigate, for the first time, how to *pseudo-label* the unlabeled mmWave PCs, and *convert (or translate)* a LiDAR dataset into its mmWave counterpart. Both the pseudo-labeled and converted mmWave data then greatly expand the original mmWave dataset for mmWave HPE model training.

Previous attempts to expand mmWave datasets are based on extracting and converting the skeletons from video datasets to mmWave PCs [9, 13]. However, these methods focus on expanding human actions, whereas the PCs follow a similar distribution as the original mmWave data without diversifying their attributes. Some other methods augment mmWave PCs using LiDAR PCs as supervision signals [6, 14, 25, 27]. While commendable, they require mmWave-LiDAR pairing and their simultaneous joint labeling in the dataset. Such multi-modal data is scarce, inconvenient to collect, and costly to label in practice, hence limiting their deployability.

In this work, we make the following contributions:

- *A Trained Pseudo-label Estimator for the Unlabeled mmWave Point Cloud:* To turn unlabeled mmWave PCs into training data, we propose a simple yet effective pseudo-label estimator trained with a supervised loss on an mmWave dataset, and an unsupervised temporal consistency loss (UTCL) on unlabeled mmWave data. Our UTCL improves the estimation robustness by enforcing temporal consistency in the predicted pseudo-labels.
- *A Closed-Form Converter to Translate LiDAR Dataset to mmWave Point Cloud:* We propose a closed-form PC converter, an approach independent of pseudo-label estimator that translates a LiDAR dataset to its mmWave counterpart. The converter uses a flow-based point filtering (FPF) algorithm to realistically capture the motion detection mechanism of mmWave PCs where moving body parts (high-flow points) are more likely to be detected than static ones (low-flow points). By integrating FPF with traditional PC augmentation techniques (*e.g.*, noise injection, random sampling), EMDUL effectively translates an input LiDAR dataset into its mmWave counterpart with realistic PC attributes.
- *EMDUL, A Novel Pipeline to Expand an mmWave Dataset:* We propose EMDUL, a novel pipeline to greatly **e**xpand an **m**mWave **d**ataset with **u**nlabeled mmWave data and **L**iDAR dataset. EMDUL first uses the closed-form converter to translate LiDAR dataset to its mmWave counterpart. Using the converted and existing mmWave datasets, it subsequently trains the pseudo-label estimator. The trained estimator is then used to annotate unlabeled mmWave PC. Expanded with the converted and pseudo-labeled data, the original mmWave dataset is greatly enhanced in terms of volume and diversity. This dataset is then used to train the final HPE inference models.

To validate EMDUL, we conduct extensive experiments on commonly used mmWave datasets, MM-Fi [37] and mmBody [5] with portions designated as unlabeled data, and LiDAR datasets LiDARHuman26M [21] and HmPEAR [22]. Our results show that EMDUL significantly improves model performance and generalization compared to training solely on the original mmWave dataset. In our generalization study, EMDUL achieves a substantial 15.1% error reduction when trained and tested on MM-Fi (in-domain study), and a 18.9% reduction when trained on mmBody and tested on MM-Fi (out-of-domain study) using unlabeled data and HmPEAR for expansion.

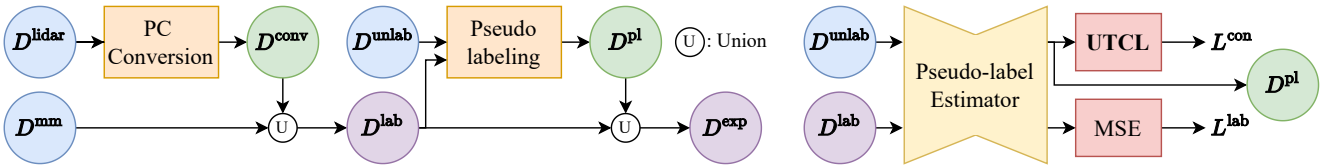
2. Related Works

2.1. mmWave-based Human Pose Estimation

In recent years, mmWave radar has attracted increasing attention for HPE due to its low deployment cost, 3D data capture, privacy-friendly nature, and robustness under various lighting conditions. While some methods directly use raw 4D radar tensor as input [10, 15, 20, 38, 40], the high computational cost and unavailability on certain hardware platforms limit their application scenarios. Therefore, mainstream mmWave HPE approaches often use processed 3D point clouds (PCs) [1, 5, 7, 12, 31, 32], which reduces the computational demand and ensures broader device compatibility. However, existing datasets for mmWave PC-based HPE [2, 5, 8, 37] are scarce and lack diversity in PC attributes and human poses, leading to poor generalization on unseen testing scenarios.

2.2. Data Expansion or Augmentation for mmWave Datasets

To address the data scarcity issue in mmWave-based HPE, various data expansion or augmentation strategies have been investigated. Current data *expansion* methods for HPE or related human sensing tasks are typically based on extracting additional skeletons from video data and converting them to PCs by leveraging statistical properties of the original mmWave dataset [9, 13]. Although these methods increase pose diversity, the generated PCs still follow a similar distribution to the original dataset. Compared to mmWave, LiDAR HPE datasets also contain 3D PCs and generally capture a wider range of human poses [21, 22, 29],



(a) The overall pipeline of EMDUL (b) The pseudo-labeling process

Figure 2. The overview of EMDUL integrating both PC conversion and pseudo-labeling modules.

making them an ideal resource for expanding mmWave data in both PC and pose diversity. However, due to the distinct attributes between LiDAR and mmWave PCs, models trained exclusively on LiDAR data exhibit poor generalization when applied to mmWave data. Existing *augmentation* techniques developed for other tasks address this issue by using LiDAR PCs as a supervision signal to improve mmWave PC density [6, 14, 25, 27]. While effective, these methods rely on co-labeled mmWave-LiDAR pairs, which are challenging to acquire in practice. Our proposed method, in contrast, efficiently leverages an independent LiDAR dataset to diversify both PCs and human poses without requiring co-labeled data.

2.3. Semi-Supervised Learning Approaches

Semi-supervised learning trains a model using limited labeled data alongside a large amount of unlabeled data. Most existing semi-supervised learning techniques fall into two categories: (1) *pseudo-labeling-based methods* [19, 28, 35], which predict pseudo-labels for unlabeled data, and (2) *consistency-based methods* [4, 18, 30, 33, 34], which apply different augmentations to the same input and enforce output similarity. Based on these techniques, various methods have been proposed for image-based HPE [16, 26, 36], but they cannot be directly applied to mmWave HPE due to differences in input modality and output formats. Our EMDUL incorporates a pseudo-labeling method specifically designed for mmWave HPE, effectively utilizing unlabeled mmWave data for dataset expansion.

3. Problem Formulation and EMDUL Overview

3.1. Problem Formulation

Let D denote an HPE dataset comprising one or multiple sequences of PC-skeleton pairs (P, S) with variable lengths. At each timestep $t \geq 0$, the input for an HPE model is a sequence of T continuous PCs $\{P_{t-T+1}, ..., P_t\}$. Each PC $P_j \in \mathbb{R}^{M_j \times 3}$ consists of M_j points represented by their Cartesian coordinates. The corresponding output is the human joint coordinates $\hat{S}_t \in \mathbb{R}^{J \times 3}$ at time t, where J is the number of joints in the skeleton. Our objective is to expand an mmWave dataset D^{mm} in terms of volume and diversity.

While mmWave PCs can include features like Doppler speed, these are often hardware-dependent and can limit a model's generalization ability. We therefore exclude them. For HPE models that require a per-point feature (*e.g.* P4Transformer [11]), we use only the point's height, which is independent of specific hardware.

3.2. EMDUL Overview

We propose EMDUL, a novel approach to expand an mmWave HPE dataset D^{mm} using two readily available data sources: unlabeled mmWave data D^{unlab} and an annotated LiDAR dataset D^{lidar}. EMDUL consists of two independent modules:

- *Pseudo-labeling of unlabeled data*, which generates pseudo-labels for D^{unlab} to form D^{pl}.
- *PC conversion of LiDAR datasets*, which translates PCs in D^{lidar} into its mmWave counterpart to form D^{conv}.

The complete pipeline of EMDUL is displayed in Fig. 2(a). First, D^{lidar} is translated into D^{conv} using the PC conversion module, which is then combined with D^{mm} to form $D^{\text{lab}} = D^{\text{mm}} \cup D^{\text{conv}}$. Next, the pseudo-labeling module processes D^{lab} and D^{unlab} to generate pseudo-labels for D^{unlab}, yielding D^{pl}. The inclusion of D^{conv} in the training data substantially improves the quality of generated pseudo-labels. The final expanded dataset is the union: $D^{\text{exp}} = D^{\text{lab}} \cup D^{\text{pl}}$.

3.3. Training on Expanded Dataset

The inference HPE model θ^{infer} is trained from scratch on the expanded dataset D^{exp} with a mean-squared error (MSE) loss. To maximize diversity, D^{conv} is re-generated at each epoch with a new random seed. Concurrently, the pseudo-estimator θ^{pl} is trained alongside θ^{infer}. During each epoch, θ^{pl} is updated first and used to generate a new D^{pl} from D^{unlab}. θ^{infer} is then trained on the updated $D^{\text{exp}} = D^{\text{lab}} \cup D^{\text{pl}}$. This iterative refinement allows θ^{infer} to incrementally extract meaningful information from D^{unlab}.

4. Pseudo-labeling of Unlabeled mmWave Data

To utilize unlabeled mmWave data for dataset expansion, we employ a simple but effective pseudo-labeling estimator. As illustrated in Fig. 2(b), the estimator θ^{pl} is trained on labeled data D^{lab} using an MSE loss L^{lab}, as well as simultaneously on unlabeled data D^{unlab} using our novel Unsuper-

vised Temporal Consistency Loss (UTCL) L^{con} (Sec. 4.1). The θ^{pl} generates pseudo-labels for D^{unlab} to form a new dataset D^{pl}.

4.1. Unsupervised Temporal Consistency Loss (UTCL)

Predicting skeletons independently at each timestep can lead to temporal inconsistency. To enhance pseudo-label reliability, we propose UTCL, motivated by a key physical insight from mmWave sensing: joints far from any detected points are likely static, whereas those embedded within the PC are likely in motion. This arises from the *motion detection mechanism* (Fig. 3) of mmWave radar: due to the reliance on the Doppler effect during PC formation, mmWave radars detect moving targets more easily than static ones. UTCL enforces this physical prior by penalizing predictions that violate it. It consists of two complementary components: a Dynamic Consistency Loss (DCL) and a Static Consistency Loss (SCL).

Given a PC P_t, its predicted skeleton $\hat{S}_t$, and the skeleton flow $\hat{F}_t^S = \hat{S}_t - \hat{S}_{t-1}$, DCL encourages joints that are likely moving to have a non-zero flow. We first identify the set of "dynamic" joints, F_t^{dyn}, as those whose distance to the nearest point in P_t is less than a threshold μ:

$$F_t^{\text{dyn}} = \{\hat{F}_t^S[j] : \min_i \|\hat{S}_t[j] - P_t[i]\|_2 < \mu\}. \quad (1)$$

Then, L^{dyn} penalizes these joints if their flow magnitude is below a flow threshold η, effectively encouraging motion:

$$L^{\text{dyn}} = \frac{1}{|F_t^{\text{dyn}}|} \sum_{k=1}^{|F_t^{\text{dyn}}|} \max(0, \eta - \|F_t^{\text{dyn}}[k]\|_2). \quad (2)$$

Similarly, SCL encourages joints that are likely static to have zero flow:

$$F_t^{\text{sta}} = \{\hat{F}_t^S[j] : \min_i \|\hat{S}_t[j] - P_t[i]\|_2 > \rho\}, \quad (3)$$

$$L^{\text{sta}} = \frac{1}{|F_t^{\text{sta}}|} \sum_{k=1}^{|F_t^{\text{sta}}|} \|F_t^{\text{sta}}[k]\|_2, \quad (4)$$

where ρ is the threshold for static joints.

The final UTCL is the sum of DCL and SCL: $L^{\text{con}} = L^{\text{dyn}} + L^{\text{sta}}$.

4.2. Training of Pseudo-label Estimator

In each training step, we sample one instance from D^{lab} and another from D^{unlab}. The overall training loss combines the supervised loss L^{lab} and UTCL L^{con} as a weighted sum:

$$L = L^{\text{lab}} + \lambda^{\text{con}} L^{\text{con}}, \quad (5)$$

where λ^{con} is the weighting parameter.

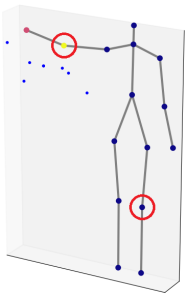
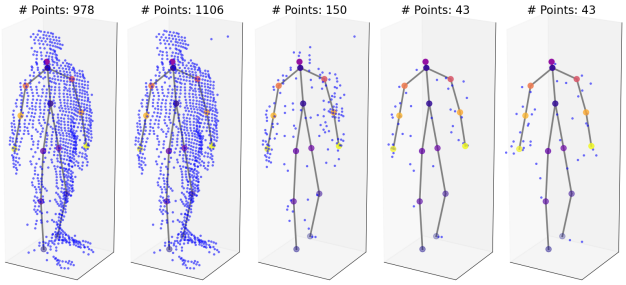


Figure 3. Illustration of the motion-detection mechanism in mmWave radar using an MM-Fi sample. Joints with high flow (yellow) lie close to detected points, while low-flow joints (dark blue) have no nearby points.



(a) Original (b) After NPA (c) After FPF (d) After RS (e) After NI

Figure 4. Step-by-step visualization of the point-cloud (PC) conversion pipeline Blue joints have lower flow magnitudes and yellow joints higher ones.

5. Converting LiDAR Datasets to mmWave Point Clouds

In addition to pseudo-labeling unlabeled mmWave data, we also convert a LiDAR dataset into its mmWave counterpart to expand the original mmWave dataset.

5.1. Closed-Form PC Conversion

Our closed-form PC conversion pipeline is a sequence of augmentations designed to simulate different attributes of mmWave PCs:

- *Noisy point addition (NPA)*: Adds a fixed number of noise points to simulate non-human environmental objects detected by an mmWave radar.
- *Flow-based Point Filtering (FPF)*: Our proposed algorithm (detailed in Sec. 5.2) that simulates the motion detection mechanism as described in Sec. 4.1.
- *Random sampling (RS)*: Reduces point density by randomly sampling a fraction of points, mimicking sparser mmWave PCs.
- *Noise injection (NI)*: Injects random noise into each point's coordinates to simulate lower spatial resolution.

The augmentations are sequentially applied in the order:

NPA → FPF → RS → NI. The progressive transformation after each step is visualized in Fig. 4, which clearly illustrates their impacts on PC attributes.

5.2. Flow-based Point Filtering (FPF)

Flow refers to the temporal displacement of points or joints. FPF simulates the motion detection mechanism of mmWave radar by interpolating PC flow based on skeleton flow, and then filtering points with lower flow magnitude with higher probability.

Given two consecutive LiDAR PCs, P_{t-1} and P_t, and their ground-truth skeletons, S_{t-1} and S_t, we aim to estimate a 3D flow vector $F_t^P[i]$ for each point $P_t[i]$ in the current PC. To create a stable flow field for interpolation, we establish boundary constraints by extending the skeletons S_{t-1} and S_t to S'_{t-1} and S'_t by adding the eight vertices of the axis-aligned bounding cube that encloses both the joints and PCs at $t-1$ and t. We define the extended skeleton flow $F_t'^S$ as the displacement of these $J+8$ points $S'_t - S'_{t-1}$. Notably, the eight static vertices naturally have zero flow.

We then interpolate the flow for each point $P_t[i]$ as a linear combination of $F_t'^S$ using inverse distance weighting. The normalized weights $\tilde{w}_t[i,j]$ for each point i relative to each of the $J+8$ extended joints j is:

$$w_t[i,j] = \frac{1}{\|P_t[i] - S'_t[j]\|_2 + \epsilon}, \tag{6}$$

$$\tilde{w}_t[i,j] = \frac{w_t[i,j]}{\sum_{k=1}^{J+8} w_t[i,k]}, \tag{7}$$

where $\epsilon = 10^{-6}$ is used to prevent division by zero. The interpolated PC flow F_t^P is then:

$$F_t^P = \tilde{w}_t F_t'^S. \tag{8}$$

This ensures that the flow of each point is most similar to its nearest joints or boundary vertices.

Finally, we simulate the motion detection mechanism by filtering P_t using the interpolated flow F_t^P. Points in P_t with lower flow magnitude are discarded with a higher probability to form the resultant PC P_t^{conv}:

$$\mathcal{P}(P_t[i] \in P_t^{\text{conv}}) = \min\left(\frac{\|F_t^P[i]\|_2}{\upsilon_t}, 1\right), \tag{9}$$

where υ_t is the flow threshold sampled from a uniform distribution $\mathrm{U}[\gamma, \delta]$, with γ and δ as hyperparameters.

6. Illustrative Experimental Results

In this section, the datasets used in experiments are first introduced in Sec. 6.1. We then detail the implementation in Sec. 6.2 and evaluation metrics in Sec. 6.3. Next, the experimental results compared with various HPE methods are shown in Sec. 6.4. Finally, Sec. 6.5 presents ablation studies and additional analysis on EMDUL.

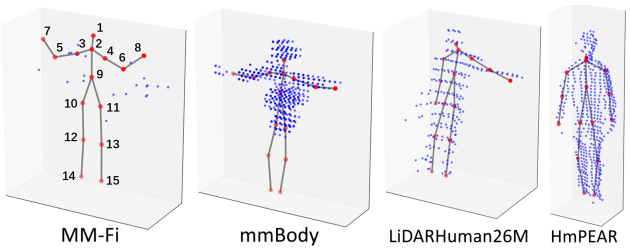


Figure 5. Sample point clouds from different mmWave and LiDAR HPE datasets and the standardized 15-keypoint skeleton structure used in this paper.

6.1. Datasets

We evaluate EMDUL on two mmWave datasets:

- *mmBody (B)* [5] is a multi-modal human sensing dataset incorporating a high-end mmWave radar, offering denser PCs that capture more static body parts than standard radars. It covers 9 scenes with 200K frames. Although it includes 200 motions performed by 30 subjects, most only involve upright postures facing the radar, resulting in limited pose diversity.
- *MM-Fi (F)* [37] is a comprehensive human pose dataset that includes mmWave radar as one modality. Compared with mmBody, PCs in MM-Fi are sparser and less sensitive to static body parts. It contains 321K frames across 4 scenes, covering 27 actions by 40 subjects with similarly limited pose diversity.

We further use two LiDAR datasets:

- *LiDARHuman26M (L)* [21] is a long-range LiDAR-based human pose dataset capturing human actions at various distances, resulting in PCs with varying, sometimes low, densities. It contains 184K frames across 2 scenes and 20 actions with more diverse poses, including ones unseen in mmBody and MM-Fi (e.g., swimming and squatting).
- *HmPEAR (H)* [22] is a recent LiDAR dataset for HPE and action recognition, consisting of 250K frames from 10 scenes. It includes 40 action categories with even greater pose diversity than LiDARHuman26M.

To unify skeleton formats, we adopt a standardized structure with 15 keypoints (Fig. 5). To simulate extreme mmWave data scarcity, D^{mm} contains PCs and skeletons from 10% randomly sampled sequences of the mmWave training set, while PCs from the remaining 90% form the unlabeled dataset D^{unlab}. D^{lidar} includes the full training set of the LiDAR dataset.

6.2. Implementation

In pseudo-labeling, we use P4T [11] or SPiKE [3] as the pseudo-label estimator. Its input is a sequence of 5 PCs, each uniformly processed to 256 points through truncation or padding. The thresholds in UTCL are configured as

Table 1. Comparison with state-of-the-art mmWave HPE methods. Each model is trained with only 10% labeled data from MM-Fi (F) or mmBody (B). Depending on the setting, methods may additionally use the remaining 90% unlabeled mmWave data and/or the LiDAR dataset HmPEAR. All results are reported in centimeters (cm), and lower is better.

Method	Backbone	F → F		F → B		B → F		B → B	
		MPJPE	PA-MPJPE	MPJPE	PA-MPJPE	MPJPE	PA-MPJPE	MPJPE	PA-MPJPE
10% labeled mmWave only									
PT [39]	PT	15.88	10.62	**17.75**	**14.03**	35.22	**14.83**	13.32	9.78
mmDiff [12]	mmDiff	15.10	10.73	44.00	21.76	**28.12**	20.15	14.98	10.69
P4T [11]	P4T	12.23	7.95	20.78	16.10	33.62	15.85	11.39	8.37
SPiKE [3]	SPiKE	**11.85**	**7.92**	18.70	14.40	37.09	16.30	**10.70**	**7.86**
+ unlabeled mmWave data									
MT [34]	P4T	26.77	14.94	**19.27**	**15.27**	52.52	21.68	15.43	10.00
MT [34]	SPiKE	19.62	11.03	20.56	16.21	47.82	17.50	12.89	8.68
EMDUL-PL	P4T	**11.36**	7.38	22.90	16.13	41.47	17.98	**10.79**	8.26
EMDUL-PL	SPiKE	11.45	**7.21**	23.93	16.04	**38.67**	**17.01**	10.83	**6.99**
+ HmPEAR LiDAR data									
P4T [11]	P4T	11.02	7.55	16.08	12.93	32.13	14.93	11.14	7.17
SPiKE [3]	SPiKE	10.41	7.15	17.29	12.81	33.90	15.34	10.90	**6.86**
EMDUL-PCC	P4T	**10.21**	**7.12**	**15.22**	**11.51**	**24.24**	14.99	**10.86**	7.09
EMDUL-PCC	SPiKE	10.59	7.36	15.41	11.71	24.25	**14.59**	10.97	7.11
Full setting (+ unlabeled mmWave data + HmPEAR LiDAR data)									
MT [34]	P4T	10.37	**6.80**	16.97	12.12	31.51	14.61	11.04	7.17
MT [34]	SPiKE	12.45	7.54	16.15	12.15	32.71	15.74	11.04	7.17
EMDUL	P4T	**10.06**	7.01	**14.89**	**11.11**	24.01	14.33	11.11	**6.89**
EMDUL	SPiKE	10.40	7.23	15.11	11.36	**22.80**	**14.09**	**10.82**	7.02

$\mu = 20\,\text{cm}$, $\eta = 5\,\text{cm}$ and $\rho = 5\,\text{cm}$, with weighting parameter $\lambda^{\text{con}} = 0.01$. We train the estimator for 100 epochs using the AdamW [23] optimizer with a learning rate 10^{-4}, and the Cosine Annealing learning rate scheduler [24] with linear warmup at an initial learning rate 10^{-5}. In FPF, the flow threshold υ for FPF is sampled uniformly between $\gamma = 2\,\text{cm}$ and $\delta = 5\,\text{cm}$. The inference HPE model mirrors the estimator's network architecture, input format, and optimization configuration. Additional implementation details are provided in the supplementary material.

6.3. Evaluation Metrics

Following previous works, we employ two commonly used evaluation metrics from Human36M [17]:

- *Mean Per Joint Position Error (MPJPE)*, which calculates the average Euclidean distance error for each joint between the predicted skeleton and the ground truth.
- *Procrustes Analysis MPJPE (PA-MPJPE)*, which measures error after aligning the predicted and ground-truth skeletons using Procrustes methods, including translation, rotation, and scaling, evaluating the quality of the overall pose structure.

6.4. Comparison with the State of the Art

This section presents quantitative results of mmWave HPE models trained on EMDUL-expanded dataset compared to those with varying data conditions. In addition to evaluating our entire approach, we separately assess our pseudo-labeling for unlabeled data (EMDUL-PL) and PC conversion method for LiDAR data (EMDUL-PCC). We classify the comparison schemes into four categories based on the data used to expand D^{lab}:

- *No data expansion*, where PT [39], mmDiff [12], P4T [11], and SPiKE [3] are trained solely on D^{mm}.
- *Expansion using a LiDAR dataset*, where P4T and SPiKE are trained on both D^{mm} and D^{lidar}.
- *Expansion using unlabeled data*, where P4T and SPiKE are trained on $D^{\text{lab}} = D^{\text{mm}}$ and D^{unlab}, integrating a widely used Mean-Teacher (MT) pseudo-labeling strategy [34] adapted by us for mmWave HPE.
- *Expansion using both a LiDAR dataset and unlabeled data*, where P4T and SPiKE are trained on $D^{\text{lab}} = D^{\text{mm}} \cup D^{\text{lidar}}$ and D^{unlab} using MT for pseudo-labeling.

Let $D^{\text{train}} \rightarrow D^{\text{test}}$ denote the setting where the model is trained on dataset D^{train} and tested on D^{test}. The trained models are evaluated on both the test split of D^{mm} (in-

Table 2. Comparison with Mean Teacher (MT) pseudo-labeling when expanding MM-Fi (F) with different LiDAR datasets. P4T [11] serves as the common HPE model.

Setting		F → F		F → B	
Method	LiDAR Dataset	MPJPE	PA-MPJPE	MPJPE	PA-MPJPE
MT	H	10.37	**6.80**	16.97	12.12
EMDUL	H	**10.06**	7.01	**14.89**	**11.11**
MT	L	10.93	**6.85**	18.62	16.31
EMDUL	L	**10.05**	6.93	**16.97**	**12.82**
MT	H+L	10.60	6.84	17.04	13.53
EMDUL	H+L	**9.92**	**6.82**	**16.43**	**12.34**

Table 3. Preliminary error comparison of different point features.

Point feature	MPJPE	PA-MPJPE
Doppler	25.15	16.25
Height	**19.68**	**15.24**

domain scenario, *e.g.*, F → F) and another mmWave dataset with unseen scenarios (out-of-domain scenario, *e.g.*, F → B). All results are presented in centimeters (cm).

The results presented in Tab. 1 show that dataset expansion generally improves performance. Specifically, pseudo-labeling (EMDUL-PL) of unlabeled data achieves better fitting for in-domain data; however, it also tends to overfit because D^{unlab} and D^{mm} share similar distributions in our setting. In contrast, PC conversion (EMDUL-PCC) for a LiDAR dataset substantially improves performance in both in-domain and out-of-domain scenarios.

When integrating both components, our proposed EMDUL consistently outperforms models with no or incomplete data expansion across most settings, achieving a significant 15.1% decrease in MPJPE on F → F, and a 18.9% MPJPE reduction on B → F. Compared to the MT approach, which also expands D^{mm} with D^{lidar} and D^{unlab}, EMDUL performs better in 7 out of 8 cases. When employing SPiKE as the same HPE model, EMDUL surpasses MT by 17.5% in MPJPE on F → F, and 30.3% in MPJPE on B → F. Results in Tab. 2 show that the improvement is general across different LiDAR datasets. These validate that EMDUL significantly improves in-domain accuracy and out-of-domain generalization of models. Notably, the performance improvement is more pronounced in out-of-domain scenarios, indicating that EMDUL effectively enhances data diversity.

6.5. Ablation Studies

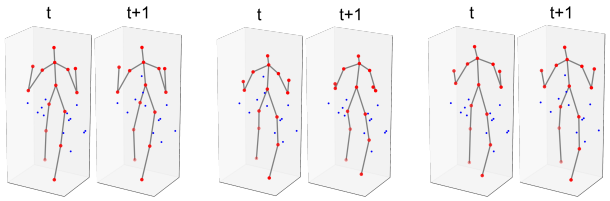
In this section, we conduct ablation studies and other analyses under the setting of training on MM-Fi + HmPEAR and testing on MM-Fi or mmBody. All methods use P4T as the HPE model. Results are presented in centimeters (cm). We assess performance in out-of-domain evaluations (F → B) in ablation studies, as our primary goal is to improve model generalization rather than in-domain fitting.

Table 4. Ablation study on pseudo-labeling of unlabeled data.

Components			F → B	
L^{lab}	L^{dyn}	L^{sta}	MPJPE	PA-MPJPE
			15.22	11.51
✓			15.12	11.44
✓	✓		15.35	11.73
✓		✓	15.53	11.80
✓	✓	✓	**14.89**	**11.11**

Table 5. Ablation study on PC conversion of a LiDAR dataset.

Components				F → B	
NPA	FPF	RS	NI	MPJPE	PA-MPJPE
				15.85	12.23
✓				15.80	12.23
	✓			15.63	11.80
✓	✓			15.49	11.74
✓	✓	✓		15.47	11.56
✓		✓	✓	15.54	11.75
✓	✓	✓	✓	**14.89**	**11.11**



(a) GT (b) No UTCL (c) Ours

Figure 6. Comparison of pseudo-labels generated with and without UTCL. (a) Two consecutive ground-truth skeletons in D^{unlab}. (b) Pseudo-labels generated without using UTCL, (c) Pseudo-labels generated by EMDUL using UTCL.

Preliminary study on point features: We conduct a preliminary study to compare two point-level features, Doppler speed and height, under F → B. Each experiment is repeated with five different random seeds, and the results are averaged. As shown in Tab. 3, using height achieves the lower error, indicating better cross-dataset generalization.

Ablation on pseudo-labeling: We investigate the impact of our pseudo-labeling strategy to utilize unlabeled mmWave data for dataset expansion. Results in Tab. 4 show that training the pseudo-label estimator on both D^{lab} (with L^{lab}) and D^{unlab} (with UTCL: DCL L^{dyn} and SCL L^{sta})

Table 6. Ablation study on hyperparameters in UTCL and FPF under F→B.

UTCL Hyperparameters									FPF Hyperparameters					
μ	MPJPE	PA-MPJPE	η	MPJPE	PA-MPJPE	ρ	MPJPE	PA-MPJPE	γ	MPJPE	PA-MPJPE	δ	MPJPE	PA-MPJPE
1	15.37	11.67	1	15.28	11.37	10	15.67	11.90	1	15.43	11.38	2	15.25	11.60
5	**14.89**	**11.11**	5	**14.89**	**11.11**	20	**14.89**	**11.11**	2	**14.89**	**11.11**	5	**14.89**	**11.11**
10	15.52	11.86	10	15.67	11.74	50	14.93	11.12	5	15.47	11.78	10	15.58	11.48

Table 7. Ratio of PCs classified as mmWave in a binary classification task distinguishing between and mmWave and LiDAR data.

Dataset	D^{mm}	D^{lidar}	D^{conv} w/o FPF	D^{conv} w/ FPF
Ratio (%)	99.61	2.66	43.06	**60.46**

Table 8. Performance under different ratios of labeled mmWave data (D^{lab} / MM-Fi).

Setting		F → F		F → B	
Method	Labeled Ratio	MPJPE	PA-MPJPE	MPJPE	PA-MPJPE
MT	1%	18.40	13.20	19.94	12.23
EMDUL		**14.77**	**10.37**	**15.95**	**11.46**
MT	10%	10.37	**6.80**	16.97	12.12
EMDUL		**10.06**	7.01	**14.89**	**11.11**
MT	50%	8.25	**5.63**	17.02	12.15
EMDUL		**8.15**	5.68	**15.61**	**11.46**

significantly reduces error. Although the individual UTCL components, DCL and SCL, do not independently improve out-of-domain generalization, their combination yields substantial gains with only marginal in-domain error increase. This is because DCL or SCL alone biases predictions toward overly dynamic or static outcomes, while their combination maintains a better balance.

Ablation on PC conversion: We evaluate the effectiveness of our PC conversion pipeline in transforming a LiDAR dataset into its mmWave counterpart. As shown in Tab. 5, each augmentation technique incrementally improves performance. However, the most substantial gain is observed when our proposed FPF is included.

Effects of hyperparameters: We investigate the effects of key hyperparameters in UTCL (μ, η, and ρ) and FPF (γ and δ), as shown in Tab. 6. Each hyparameter is adjusted while the others are fixed. The model achieves optimal results with our selected hyperparameters under the F → B setting.

Visualization of pseudo-labels: We visualize the pseudo-labels generated with or without our UTCL in Fig. 6. As illustrated, predictions without UTCL exhibit noticeable temporal inconsistencies, particularly in the hand regions. In contrast, our UTCL effectively enforces temporal consistency, resulting in smoother and more accurate pseudo-labels that closely align with the ground truth.

Quantitative Analysis of converted PCs: To quantitatively assess how realistic the converted LiDAR PCs are, we train a binary classifier based on P4T [11] to distinguish between mmWave and LiDAR PCs. As reported in Tab. 7, 60.46% of converted LiDAR PCs are classified as mmWave data when FPF is employed, compared to only 43.06% without FPF, demonstrating that our generated mmWave PCs are realistic and highlighting the importance of simulating the motion detection mechanism using FPF.

Results with different labeled data availability: Tab. 8 examines the performance of EMDUL under varying ratios of D^{mm} over the entire MM-Fi training set. The results indicate that EMDUL consistently outperforms the MT approach across all data ratios. Notably, the most substantial error reduction is observed at 1% labeled data, demonstrating the effectiveness of EMDUL in addressing data scarcity.

7. Conclusion

We propose EMDUL, a novel mmWave training approach to address data scarcity and limited diversity in mmWave HPE training by effectively utilizing unlabeled mmWave data and annotated LiDAR dataset. EMDUL expands an mmWave dataset through two key components: a pseudo-label estimator trained with an unsupervised temporal consistency loss to generate reliable pseudo-labels for unlabeled mmWave data, and a PC conversion method to convert LiDAR PCs into its mmWave counterparts by simulating mmWave PC attributes, including flow-based point filtering to simulate motion detection. Augmented with pseudo-labeled mmWave data and both converted LiDAR dataset, the original mmWave dataset is substantially expanded in volume and diversity. Experiments on multiple mmWave and LiDAR datasets show that models trained with EMDUL substantially outperform those trained on the original mmWave dataset alone, achieving superior accuracy and generalization across both in-domain and out-of-domain scenarios.

8. Acknowledgement

This work was supported, in part, by Research Grants Council Collaborative Research Fund (under grant number C1045-23G) and RGC-General Research Fund (under grant number 16201625) of Hong Kong.

References

[1] Sizhe An and Umit Y. Ogras. Fast and scalable human pose estimation using mmWave point cloud. *Proceedings of the 59th ACM/IEEE Design Automation Conference*, pages 889–894, 2022. 2

[2] Sizhe An, Yin Li, and Umit Ogras. mRI: Multi-modal 3D Human Pose Estimation Dataset using mmWave, RGB-D, and Inertial Sensors, 2022. 2

[3] Irene Ballester, Ondřej Peterka, and Martin Kampel. SPiKE: 3D Human Pose from Point Cloud Sequences, 2024. 5, 6

[4] David Berthelot, Nicholas Carlini, Ian Goodfellow, Avital Oliver, Nicolas Papernot, and Colin Raffel. MixMatch: A holistic approach to semi-supervised learning. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, number 454, pages 5049–5059. Curran Associates Inc., Red Hook, NY, USA, 2019. 3

[5] Anjun Chen, Xiangyu Wang, Shaohao Zhu, Yanxu Li, Jiming Chen, and Qi Ye. mmBody Benchmark: 3D Body Reconstruction Dataset and Analysis for Millimeter Wave Radar, 2023. 2, 5

[6] Yuwei Cheng, Jingran Su, Mengxin Jiang, and Yimin Liu. A Novel Radar Point Cloud Generation Method for Robot Environment Perception. *IEEE Transactions on Robotics*, 38(6):3754–3773, 2022. 2, 3

[7] Han Cui and Naim Dahnoun. Real-Time Short-Range Human Posture Estimation Using mmWave Radars and Neural Networks. *IEEE Sensors Journal*, 22(1):535–543, 2022. 2

[8] Han Cui, Shu Zhong, Jiacheng Wu, Zichao Shen, Naim Dahnoun, and Yiren Zhao. Milipoint: A point cloud dataset for mmwave radar. In *Advances in Neural Information Processing Systems*, pages 62713–62726, 2023. 2

[9] Kaikai Deng, Dong Zhao, Qiaoyue Han, Zihan Zhang, Shuyue Wang, Anfu Zhou, and Huadong Ma. Midas: Generating mmWave Radar Data from Videos for Training Pervasive and Privacy-preserving Human Sensing Tasks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 7(1): 9:1–9:26, 2023. 2

[10] Wen Ding, Zhongping Cao, Jianxiong Zhang, Rihui Chen, Xuemei Guo, and Guoli Wang. Radar-Based 3D Human Skeleton Estimation by Kinematic Constrained Learning. *IEEE Sensors Journal*, 21(20):23174–23184, 2021. 2

[11] Hehe Fan, Yi Yang, and Mohan Kankanhalli. Point 4D Transformer Networks for Spatio-Temporal Modeling in Point Cloud Videos. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14199–14208, 2021. 1, 3, 5, 6, 7, 8

[12] Junqiao Fan, Jianfei Yang, Yuecong Xu, and Lihua Xie. Diffusion Model is a Good Pose Estimator from 3D RF-Vision, 2024. 2, 6

[13] Yuxin Fan, Yong Wang, Hang Zheng, and Zhiguo Shi. Video2mmPoint: Synthesizing mmWave Point Cloud Data From Videos for Gait Recognition. *IEEE Sensors Journal*, 25(1):773–782, 2025. 2

[14] Zeyu Han, Junkai Jiang, Xiaokang Ding, Jiahao Wang, Qingwen Meng, Shaobing Xu, Lei He, and Jianqiang Wang. DenserRadar: A 4D Millimeter-Wave Radar Point Cloud Detector Based on Dense LiDAR Point Clouds. In *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*, pages 930–936, 2024. 2, 3

[15] Yuan-Hao Ho, Jen-Hao Cheng, Sheng Yao Kuan, Zhongyu Jiang, Wenhao Chai, Hsiang-Wei Huang, Chih-Lung Lin, and Jenq-Neng Hwang. RT-Pose: A 4D Radar Tensor-Based 3D Human Pose Estimation and Localization Benchmark. In *Computer Vision – ECCV 2024*, pages 107–125, Cham, 2025. Springer Nature Switzerland. 2

[16] Linzhi Huang, Yulong Li, Hongbo Tian, Yue Yang, Xiangang Li, Weihong Deng, and Jieping Ye. Semi-Supervised 2D Human Pose Estimation Driven by Position Inconsistency Pseudo Label Correction Module. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 693–703, 2023. 3

[17] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339, 2014. 6

[18] S. Laine and Timo Aila. Temporal Ensembling for Semi-Supervised Learning. *ArXiv*, 2016. 3

[19] Dong-Hyun Lee. Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. 2013. 3

[20] Shih-Po Lee, Niraj Prakash Kini, Wen-Hsiao Peng, Ching-Wen Ma, and Jenq-Neng Hwang. HuPR: A Benchmark for Human Pose Estimation Using Millimeter Wave Radar. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5704–5713, 2023. 2

[21] Jialian Li, Jingyi Zhang, Zhiyong Wang, Siqi Shen, Chenglu Wen, Yuexin Ma, Lan Xu, Jingyi Yu, and Cheng Wang. LiDARCap: Long-range Markerless 3D Human Motion Capture with LiDAR Point Clouds. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20470–20480, 2022. 2, 5

[22] Yitai Lin, Zhijie Wei, Wanfa Zhang, Xiping Lin, Yudi Dai, Chenglu Wen, Siqi Shen, Lan Xu, and Cheng Wang. HmPEAR: A Dataset for Human Pose Estimation and Action Recognition. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 2069–2078, New York, NY, USA, 2024. Association for Computing Machinery. 2, 5

[23] I. Loshchilov and F. Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2017. 6

[24] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic Gradient Descent with Warm Restarts, 2017. 6

[25] Kai Luan, Chenghao Shi, Neng Wang, Yuwei Cheng, Huimin Lu, and Xieyuanli Chen. Diffusion-Based Point Cloud Super-Resolution for mmWave Radar Data. In *2024*

IEEE International Conference on Robotics and Automation (ICRA), pages 11171–11177, 2024. 2, 3

[26] Dario Pavllo, Christoph Feichtenhofer, David Grangier, and Michael Auli. 3D Human Pose Estimation in Video With Temporal Convolutions and Semi-Supervised Training. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7745–7754, 2019. 3

[27] Akarsh Prabhakara, Tao Jin, Arnav Das, Gantavya Bhatt, Lilly Kumari, Elahe Soltanaghai, Jeff Bilmes, Swarun Kumar, and Anthony Rowe. High Resolution Point Clouds from mmWave Radar. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4135–4142, 2023. 2, 3

[28] Ilija Radosavovic, Piotr Dollár, Ross Girshick, Georgia Gkioxari, and Kaiming He. Data Distillation: Towards Omni-Supervised Learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4119–4128, 2018. 3

[29] Yiming Ren, Xiao Han, Chengfeng Zhao, Jingya Wang, Lan Xu, Jingyi Yu, and Yuexin Ma. LiveHPS: LiDAR-Based Scene-Level Human Pose and Shape Estimation in Free Environment. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1281–1291, 2024. 2

[30] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 1171–1179, Red Hook, NY, USA, 2016. Curran Associates Inc. 3

[31] Arindam Sengupta and Siyang Cao. *mmPose-NLP* : A Natural Language Processing Approach to Precise Skeletal Pose Estimation Using mmWave Radars. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):8418–8429, 2023. 2

[32] Akash Deep Singh, Sandeep Singh Sandha, Luis Garcia, and Mani Srivastava. RadHAR: Human Activity Recognition from Point Clouds Generated through a Millimeter-wave Radar. *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, pages 51–56, 2019. 2

[33] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. FixMatch: Simplifying semi-supervised learning with consistency and confidence. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, pages 596–608, Red Hook, NY, USA, 2020. Curran Associates Inc. 3

[34] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 1195–1204, Red Hook, NY, USA, 2017. Curran Associates Inc. 3, 6, 1

[35] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V. Le. Self-Training With Noisy Student Improves ImageNet Classification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, 2020. 3

[36] Rongchang Xie, Chunyu Wang, Wenjun Zeng, and Yizhou Wang. An Empirical Study of the Collapsing Problem in Semi-Supervised 2D Human Pose Estimation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11220–11229, 2021. 3

[37] Jianfei Yang, He Huang, Yunjiao Zhou, Xinyan Chen, Yuecong Xu, Shenghai Yuan, Han Zou, Chris Xiaoxuan Lu, and Lihua Xie. MM-Fi: Multi-Modal Non-Intrusive 4D Human Dataset for Versatile Wireless Sensing, 2023. 2, 5

[38] Peijun Zhao, Chris Xiaoxuan Lu, Bing Wang, Niki Trigoni, and Andrew Markham. CubeLearn: End-to-End Learning for Human Motion Recognition From Raw mmWave Radar Signals. *IEEE Internet of Things Journal*, 10(12):10236–10249, 2023. 2

[39] Ce Zheng, Sijie Zhu, Matias Mendieta, Taojiannan Yang, Chen Chen, and Zhengming Ding. 3D Human Pose Estimation with Spatial and Temporal Transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11636–11645, 2021. 6

[40] Bing Zhu, Zixin He, Weiyi Xiong, Guanhua Ding, Jianan Liu, Tao Huang, Wei Chen, and Wei Xiang. ProbRadarM3F: mmWave Radar based Human Skeletal Pose Estimation with Probability Map Guided Multi-Format Feature Fusion, 2024. 2