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

Supplementary Material

Table 9. The entire PC conversion process in our EMDUL.

Component	Parameters	Explanation
NPA	$\sigma_1 = 2\,\text{cm}$, $p = 0.5$, $n = 32$	n points sampled from $\mathcal{N}(C, \sigma_1^2 I)$ to a portion of p of the LiDAR PCs, where C is the center of the skeleton and I is a 3×3 identity matrix.
FPF	$\gamma = 2\,\text{cm}$, $\delta = 10\,\text{cm}$	The flow threshold is sampled from U$[\gamma, \delta]$.
RS	$r_{\text{min}} = 0.125$ $r_{\text{max}} = 1.0$ $m = 128$	A fraction of $r \in U[r_{\text{min}}, r_{\text{max}}]$ points are randomly sampled from the PC if it contains at least m points.
NI	$\sigma_2 = 5\,\text{cm}$	A noise following $N(O, \sigma_2^2 I)$ is injected to each point, where O is the origin and I is a 3×3 identity matrix.

9. More Implementation Details

This section presents more implementation details of our proposed EMDUL and a comparison scheme adapted for mmWave HPE.

9.1. PC Conversion Pipeline

We specify the parameters used in our PC conversion pipeline in Tab. 9. The parameters are chosen based on empirical results on the validation set. v is re-sampled per instance.

9.2. Model Training

The input PCs to the pseudo-label estimator and the inference HPE model are first normalized by subtracting the $(\tilde{X}, \tilde{Y}, \min Z)$, where $\tilde{X}$, $\tilde{Y}$ are the medians of all X and Y coordinates of the skeleton sequence, and $\min Z$ is the minimum Z (height) coordinate among all points in the skeleton sequence. Outlier points are subsequently removed using a box filter with a range of $[-1.5\,\text{m}, 1.5\,\text{m}]$ along the X and Y axes, and $[0\,\text{m}, 2.0\,\text{m}]$ along the Z axis. After outlier removal, we apply the following augmentations during training: random rotation along the vertical axis within $[-10°, 10°]$, random scaling within $[0.9, 1.1]$, and random translation within $[-1\,\text{cm}, 1\,\text{cm}]$ along each axis. Finally, each PC is processed to 256 points through repetitive padding or truncation. In the learning rate scheduler, the duration of linear warmup is 20 epochs.

Table 10. Ablation studies on the LiDAR data size, represented by the ratio of data used in HmPEAR.

Setting		F → F		F → B	
Method	Labeled Ratio	MPJPE	PA-MPJPE	MPJPE	PA-MPJPE
MT	10%	11.84	8.13	16.29	12.14
EMDUL	10%	**10.59**	**7.56**	**15.64**	**12.07**
MT	50%	10.65	7.40	16.26	**12.05**
EMDUL	50%	**10.27**	**7.19**	**16.25**	12.55
MT	100%	10.37	**6.80**	16.97	12.12
EMDUL	100%	**10.06**	7.06	**14.89**	**11.11**

Table 11. 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
		✓		15.63	12.01
			✓	15.84	12.22
✓	✓	✓		15.47	11.56
✓	✓		✓	15.70	11.69
✓		✓	✓	15.54	11.75
	✓	✓	✓	15.37	11.58
✓	✓	✓	✓	**14.89**	**11.11**

9.3. Adapted Mean-Teacher Pseudo-Labeling

Mean-Teacher (MT) Pseudo-Labeling [34] is a commonly used semi-supervised learning method designed for classification. We adapt it to mmWave HPE as a comparison scheme. The pseudo-label estimator (mean-teacher model) is an exponential moving average of the inference HPE model, with a decay rate of 0.999. The same augmentations are applied to the inputs of both the mean-teacher model and the inference model during pseudo-labeling.

10. More Experimental Results

In this section, we show more quantitative and visualization results for EMDUL.

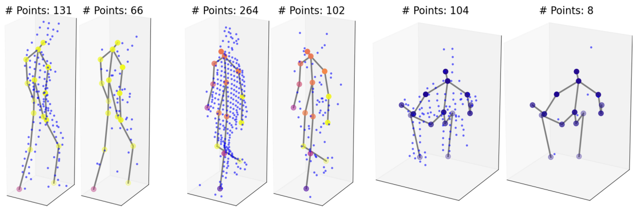


Figure 7. More visualization results of point cloud conversion. Left: original LiDAR PCs. Right: converted mmWave PCs. Joints with high flow values are yellow, while those with low flow values are blue.

10.1. Complete Ablation Study on PC Conversion

We present a more complete ablation study on each component of our PC conversion pipeline in Tab. 11. Results show that solely using each individual component brings performance gains, while removing any component from the full pipeline degrades performance. This demonstrates that each component plays a vital role in effectively converting LiDAR PCs to simulate mmWave PCs.

10.2. Results with different LiDAR data availability

To evaluate EMDUL under varying LiDAR data availability, we conduct experiments using a different ratio of data from HmPEAR. Results in Tab. 10 show that EMDUL significantly improves performance even with only 10% of HmPEAR, demonstrating its efficiency in leveraging LiDAR data.

10.3. More Visualization on PC Conversion

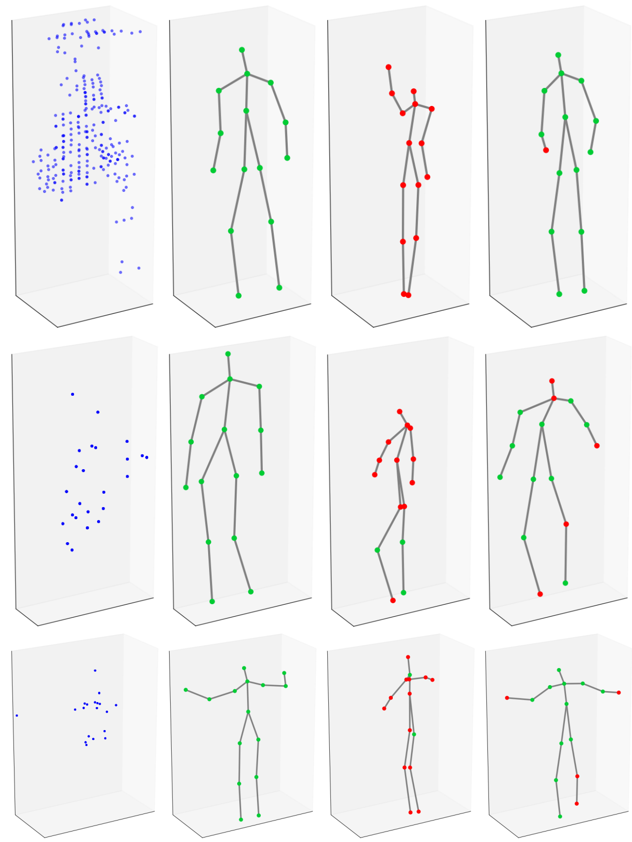
We provide more visualization results of our PC conversion pipeline in Fig. 7. It can be observed that the converted PCs exhibit increased noisiness, reduced point density, with more points gathering around fast moving joints, effectively simulating mmWave PC attributes.

10.4. More Visualizations on HPE Results

Fig. 8 shows more visualization results for EMDUL. The first row compares P4T predictions trained on MM-Fi without versus with EMDUL (augmented by HmPEAR), while subsequent rows compare models trained on mmBody [5] expanded with LiDARHuman26M [21]. It is clearly shown that using EMDUL leads to consistently higher performance, even on sparse and noisy PCs.

11. Limitation and Future Work

While EMDUL significantly improves performance by expanding mmWave datasets with unlabeled data and LiDAR datasets, it has certain limitations that pave the way for future research. First, the PC conversion pipeline relies on empirical parameter settings, which may not be optimal for all scenarios. Future work could explore adaptive or learnable conversion methods for better simulating mmWave PC attributes from different LiDAR datasets. Second, while UCTL effectively encourages temporal consistency, it may not fully capture complex motion patterns. Future research could investigate more sophisticated temporal modeling techniques, further refining the pseudo-label quality. Lastly, EMDUL currently focuses on single-person HPE; extending it to multi-person scenarios would be a valuable direction, which includes addressing challenges such as occlusion and interaction between multiple subjects.



(a) Input PC (b) GT (c) P4T (d) Ours

Figure 8. Visualization results on MM-Fi $\rightarrow$ mmBody (first row) and mmBody $\rightarrow$ MM-Fi (second and third rows). Joints with high error (> 10 cm) are colored red while others are colored green.