mmPoint: Dense Human Point Cloud Generation from mmWave

Qian Xie[1]
qian.xie@cs.ox.ac.uk
Qianyi Deng[1]
qianyi.deng@wolfson.ox.ac.uk
Ta-Ying Cheng[1]
ta-ying.cheng@cs.ox.ac.uk
Peijun Zhao[2]
zhaoymn@mit.edu
Amir Patel[3]
amir.patel@uct.ac.za
Niki Trigoni[1]
niki.trigoni@cs.ox.ac.uk
Andrew Markham[1]
andrew.markham@cs.ox.ac.uk

[1] University of Oxford
Oxford, UK
[2] Massachusetts Institute of Technology
Massachusetts, USA
[3] University of Cape Town
Cape Town, South Africa

Abstract

Millimeter-wave (mmWave) radars have emerged as a promising technology for sensing humans in diverse environments, owing to their ability to easily obtain 3D information in the form of point clouds. However, mmWave point clouds are typically characterized by sparsity and irregularity, which may limit their potential for certain applications. To address this issue, we propose mmPoint, the first model capable of generating dense human point clouds from mmWave radar signals. Specifically, mmPoint takes a single radar frame of a human as input and generates a dense point cloud that accurately reflects the shape of the detected human as output. The proposed model consists of a novel Encoder-Decoder architecture that utilizes a Multi-Modal Encoder (MME) to extract features from both the radar signal and a point cloud template. A Multi-Resolution Decoder (MRD) is then utilized to gradually infer a dense point cloud in a three-step fashion, with a Lift-and-Deform Module (LDM) employed at each step to increase the number of points and deform the point cloud based on the radar feature. Experimental results demonstrate that mmPoint achieves excellent performance on dense point cloud generation from mmWave radar signals. Code and dataset are available at https://github.com/NUAAXQ/mmPoint.

1 Introduction

The proliferation of human-centered intelligent applications such as surveillance [36], smart control [34], AR/VR [26, 41], and fitness tracking [28] has created a pressing need for robust and non-intrusive sensing technologies. While cameras have been the predominant sensing modality for such applications [8], they are vulnerable to harsh environments (e.g., poor illumination, smoke, and fog), privacy concerns, and intrusive user experiences. In contrast, wireless radio frequency (RF) signals have emerged as an alternative that is impervious to adverse environments, privacy-preserving, and non-intrusive to users [29].

In particular, single-chip millimeter wave (mmWave) radar is a low-cost sensor that can provide 3D information (i.e., 3D point cloud) of detected targets, making it a promising sensing modality for human-centered intelligent applications. However, the sparsity of mmWave point clouds has limited the accuracy of mmWave radar for some high-precision applications [3, 18, 21, 24, 25]. The sparsity issue in mmWave point clouds arises due to the inherent limitations in angular resolution, both in the azimuth and elevation dimensions. This limitation is primarily attributed to the utilization of low-cost radars equipped with a limited number of antennas, typically on the order of 3x4. Consequently, the classical processing approach known as Constant False Alarm Rate (CFAR) operates as a peak detector, resulting in the generation of highly sparse point clouds comprising only $64 \sim 128$ points. Such sparsity poses significant challenges in terms of interpretation and is unsuitable for accurate human-related tasks.

To the best of our knowledge, there is currently no existing research on dense human point cloud generation from mmWave signals due to the above-mentioned challenges. To address this gap, we draw inspiration from point cloud generation from single images and formulate the point cloud generation problem from mmWave signals as a mmWave-conditioned point cloud deformation task. As illustrated in Fig. 1(a), our proposed method takes a template human point cloud as input and predicts the corresponding movement for each point to drive the template point cloud to transform into the target human point cloud. mmWave signals are used as control conditions to predict the corresponding movement. To reduce the learning difficulty, we further propose a multi-resolution approach in which we adopt a three-step strategy, as illustrated in Fig. 1(b). Through steps, we use fewer points in the initial steps of the decoder and gradually increase the number of points through steps. This scheme can help to relieve the problem of the slow training procedure compared to one-step methods. The three-step strategy enables our method to progressively refine and enhance the generated point clouds, leading to more accurate and densely populated human point clouds.

The key contributions are summarized as follows:

- We introduce a novel task of generating dense human point clouds from mmWave radar signals, which has significant potential for various applications but has received little attention in the research community.
- To address this task, we formulate the point cloud generation problem as a point cloud deformation problem, and propose mmPoint, the first model that focuses on this task to the best of our knowledge.
- We create a new dataset of training pairs, consisting of mmWave signals and their corresponding dense human point clouds, to facilitate research in this direction.

2 Related Work

Point Cloud Generation. Pivotal work in deep learning-based methods on point cloud can be dated back as early as 2016 [19, 20] when per-point local features are introduced to

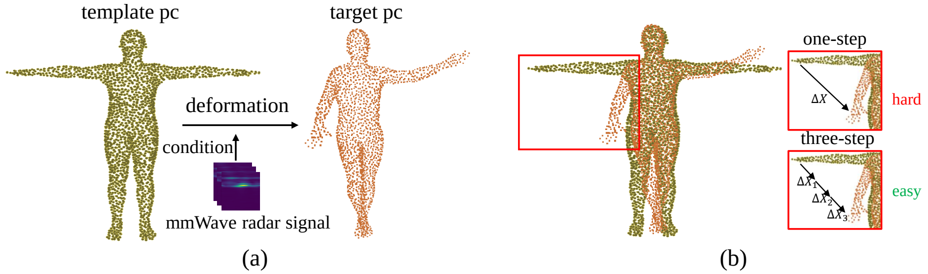

Figure 1: (a) We formulate the problem of point cloud generation from mmWvae radar signals as the problem of point cloud deformation with the condition of mmWave radar signal. (b) We design a three-step deformation strategy to gradually shift the points of the template to the target positions. The learning difficulty of the proposed three-step strategy is lower than the traditional one-step strategy.

modern MLP architectures. These pioneering architectures led to numerous soon-followed works, particularly in the task of reconstructing dense point clouds. Prior literature point cloud reconstructions span across various inputs from images [12, 14, 15] to point clouds of various forms (e.g., manipulate shape from one dense point cloud to another [2]; complete full point cloud given only parts [27, 37]; generate a dense point cloud from sparse and noisy inputs [7]). The majority of the work utilizes a variation of encoder-decoder architectures [12, 38]. For example, Lin *et al.* [12] applies the latent features of an image encoder into a 2D decoder to obtain renderings at multiple viewpoints. On the other hand, some recent arts incorporated generative techniques such as diffusion models [13, 15]. Some work also focused on mining out the critical features used for point cloud reconstruction [6, 9].

While various modalities of inputs have been well explored, the emergence of millimetre-Wave radar is very recent and therefore under-researched. The direct transfer of previous methods is not possible due to large domain gaps, and the few works on this task are all in traditional non-deep-learning methods [21].

Human-oriented mmWave Processing. Owing to its better generalization towards different lightings, there exist many works using mmWave for human processing with tasks of human pose estimation [22], action and gait recognition [11, 24], human detection and tracking, [39] and so on. Chen *et al.* [5] established a benchmark in mapping sparse human point clouds from mmWave to 3D meshes. Nalci *et al.* [16] derived a pipeline to use raw Frequency Modulated Continuous Waves (FMCW) signals to recognize human actions. An *et al.* [3] derive a method to combine meta-learning and representation of multiple frames for a fast human pose estimation method. Nevertheless, all these methods require a complicated backbone owing to the noisy and sparse characteristics of raw mmWave data.

In light of this, we propose to develop the fundamental intermediate step of inflating and denoising mmWave inputs into dense and accurate point clouds. This would amortize the training cost for complex architectures on different objectives, while simultaneously bridging many already-matured networks on dense point clouds for downstream tasks.

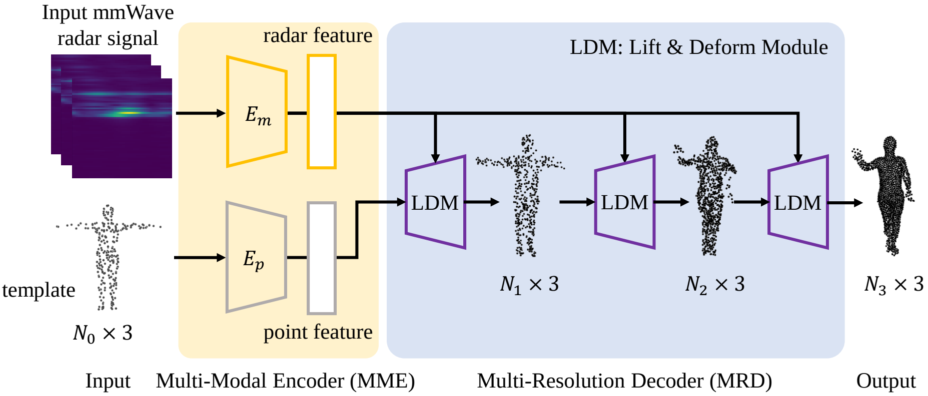

Figure 2: Overview of mmPoint. Given a single mmWave radar signal and a template human point cloud as inputs, mmPoint first transfers the inputs into representative features via MME and then generate the target point cloud via MRD.

3 Method

3.1 Problem Formulation

Given a single frame of mmWave radar signal I_m, the objective of the proposed model $\mathcal{M}_\Theta$ is to predict a dense point cloud $Q \subset \mathbb{R}^{N\times 3}$, where N is the number of points. The whole method can be formally formulated as:

$$Q = \mathcal{M}_\Theta(I_m) \tag{1}$$

In fact, it is quite hard to directly predict a new point cloud (i.e., the target point cloud) from a feature vector. However, it is much easier to predict the deviation from an existing point cloud (i.e., the template point cloud) to the target point cloud. Thus, we reformulate the above problem into a conditioned point cloud deformation problem. That is, a new point cloud can be generated by shifting the existing point cloud P with the predicted deviation $\Delta P \subset \mathbb{R}^{K\times 3}$. So the problem is now transformed to predict deviation ΔP, that is,

$$\Delta P = \mathcal{M}_\Theta(I_m, P) \tag{2}$$

3.2 System Pipeline

Given one single frame of human mmWave radar signal I_m and a standard human point cloud P as the template, our network is expected to output a dense point cloud Q of the detected person. The generated point cloud should not solely encapsulate the contours of the individual, but also possess an evenly distributed and sufficiently dense attribute for seamless applicability in further downstream processes, such as 3D pose inference and mesh reconstitution. As shown in Fig. 2, the proposed approach is in an Encoder-Decoder paradigm. Specifically, the input point cloud and signal first undergo a transformation into feature representation vectors through a Multi-Modal Encoder (MME), which is composed of a radar signal encoder and a point cloud encoder. Subsequently, a Multi-Resolution Decoder (MRD) is employed to gradually infer a dense point cloud that accurately captures the human shape.

In each step of the MRD, a novel Lift-and-Deform Module (LDM) is proposed to augment the number of points and deform the point cloud with regard to the radar feature. Upon completion of three steps, our network can output the final dense human point cloud.

3.3 mmWave Radar Signal Pre-processing

In this paper, we use the mmWave radar signals from the HuPR dataset [10], in which the authors collected over 200 sequences of radar signals using the TI IWR1843BOOST radars. We will discuss more details of this dataset in Sec. 4.1. In general, raw radar signals collected from mmWave radar devices comprise a set of frequency signals that bounce back from the targets [1]. Usually, radar signal processing methods, such as Fast Fourier Transform (FFT), need to be performed on the raw IF signals for estimating the motion information of the moving target, e.g., the range, Doppler, Angle-of-arrival (AoA). We use the same pre-processing method in [10], and finally get a complexed-valued tensor I_m with size $16 \times 64 \times 64 \times 8$, where the four dimensions refer to Doppler, Range, Azimuth AoA and Elevation AoA, respectively.

3.4 Network Architecture

3.4.1 Multi-Modal Encoder (MME)

To handle inputs from both mmWave signals and point cloud, our MME consists of two encoders: (1) mmWave encoder E_m and (2) point cloud encoder E_p, as illustrated in Fig. 2.

mmWave encoder: Following [10], we first reduce the dimension of the input mmWave signal I_m from $(16,64,64,8)$ to $(16,64,64)$ by mean-pooling along the elevation channel. Here, the mmWave signal can be seen as a 64×64 image with 16 channels. Then, we employ MNet [30], which is designed for mmWave radar signals, to extract information. After that, we can get an intermediate feature with the dimension of $(D,64,64)$, where D is the number of the output channels in MNet (we set 32 in our experiment). We finally adopt four 2D convolutional layers with strides to further aggregate information in this intermediate feature, resulting in a feature map with the size of $(512,8,8)$. A max-pooling of size $(8,8)$ is then adopted, followed by a fully-connected layer with an output size of 128. We, therefore, get the final mmWave feature vector $f_m \in \mathbb{R}^{1\times 128}$.

Point cloud encoder: Since the subsequent lift and deform operations are both local-structure related, we adopt EdgeConv [31] to act as the point cloud encoder E_p. An Edge-Conv layer, just like an MLP layer, can be used to embed points into feature space, while it can also capture local geometric structure via constructing graphs with local neighborhoods of points. With a 4-layer EdgeConv encoder $(3,32,64,128)$, a point cloud P with the size of $(N,3)$ can be transformed into a feature map of $(N,128)$.

3.4.2 Multi-Resolution Decoder (MRD)

In the decoder component, we introduce a three-step strategy to address the challenge of learning displacement vectors for individual points. However, this strategy comes at the expense of increased computational complexity due to multiple iterations. To mitigate this issue, we propose MRD that aims to reduce the computational overhead. This is achieved by initially processing a smaller subset of points and gradually increasing the point density in subsequent steps, resulting in dense point clouds with a larger number of points. Each

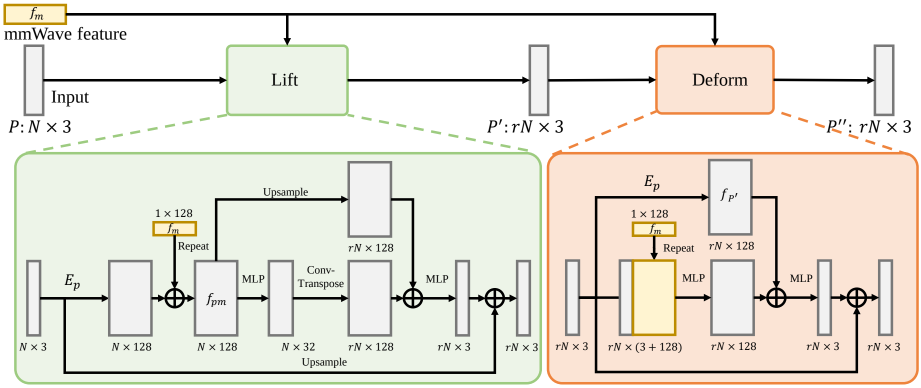

Figure 3: Architecture of the proposed Lift-and-Deform Module (LDM). Under the condition of the mmWave feature, the lift operation makes the point denser and the deform operation makes the point cloud closer to the ground truth.

decoding step comprises two fundamental operations: ***lift*** and ***deform***, which contribute to the overall decoding process. Take a point cloud $P \subset \mathbb{R}^{N\times 3}$ as input, the lift operation can increase the number of points by r times, resulting in a new point cloud $P' \subset \mathbb{R}^{rN\times 3}$. Given this lifted point cloud P', the deform operation first predicts a displacement tensor $\Delta P' \subset \mathbb{R}^{rN\times 3}$, and then outputs the deformed point $P'' = P' + \Delta P'$. r is the lifting rate and is set to 2 in our paper.

Lift operation: As shown in Fig. 3, given the input point cloud P, we first increase the number of points via a simple upsampling operation, resulting in the initial lifted points. To predict the displacement for the initial lifted points, P is also transformed into a high-dimension point feature map f_p by a point cloud encoder E_p, which is then added to the mmWave feature f_m to get the combined feature f_{pm}. Consequently, we use Upsampling and MLP-ConvTranspose operations to get two lifted features, respectively. These two lifted features are then added together, followed by another MLP layer to predict the displacement tensor. With the initial lifted point cloud and its corresponding displacement tensor, we can finally get the final lifted point cloud P'.

Deform operation: Given the lifted point cloud P' and the mmWave feature f_m, the deform operation aims to predict the deformed point cloud P'' which should be close to its corresponding ground-truth point cloud. This is also achieved by computing the displacement tensor for P' with the condition of mmWave feature f_m. Specifically, we first employ a point cloud encoder E_p to transform the input point cloud into a high-dimension feature map $f_{p'}$. Another route is directly concatenating the input point cloud and the repeated f_m, followed by an MLP layer to generate a feature map with the same size of $f_{p'}$. These two feature maps are then added together to form the displacement feature map, which is then taken as input by another MLP layer to output the final displacement tensor.

3.5 Loss Function

The training loss of our mmPoint is a joint loss that comprises two parts: 1) reconstruction loss and 2) uniform loss. Reconstruction loss penalizes the shape difference between the generated point cloud and the ground-truth point cloud. Uniform loss is used to encourage the generated points to be distributed evenly.

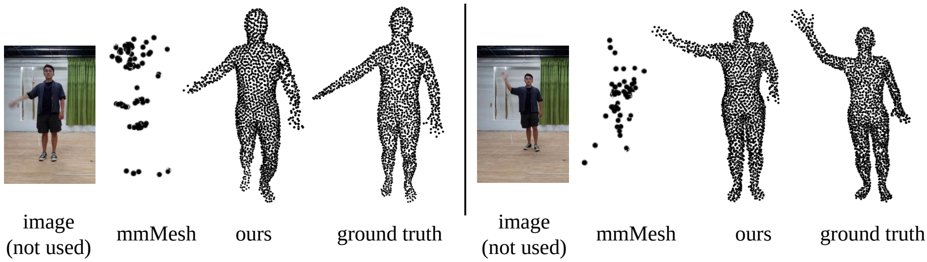

Figure 4: Our method can generate dense human point clouds from mmWave radar signal. Compared to sparse and scattered point clouds generated by mmMesh [33], our generated dense point clouds contain more geometry information and are easier to recognize the human pose. *Note that images are only displayed for reference, and not used in our network.*

Reconstruction loss. Following most of point cloud generation works, we adopt the commonly-used Chamfer Distance (CD) as our reconstruction loss $\mathcal{L}_{rec}$. That is,

$$\mathcal{L}_{rec} = \frac{1}{Q}\sum_{x\in Q}\min_{x^\star\in Q^\star}\|x-x^\star\|_2^2 + \frac{1}{Q^\star}\sum_{x^\star\in Q^\star}\min_{x\in Q}\|x^\star-x\|_2^2, \tag{3}$$

which computes the average closet point distance between the generated point cloud Q and the ground truth point cloud $Q^\star$.

Uniform loss. Minimizing $\mathcal{L}_{rec}$ can make points in Q close to their corresponding points (i.e., closest points) in $Q^\star$. However, many points in Q can correspond to the same point in $Q^\star$, resulting in a non-uniform generated point cloud. To address this issue, we adopt a uniform loss [35] $\mathcal{L}_{uniform}$ to make the generated point cloud distribute evenly. Specifically, our uniform loss is expressed as:

$$L_{uniform} = \sum_{q\in Q}\sum_{p\in\mathcal{N}(q)} \eta\left(\|p-q\|\right) w\left(\|p-q\|\right), \tag{4}$$

where $\mathcal{N}(q)$ is the point set of the k-nearest neighbors of point q, and $\|\cdot\|$ is the L2-norm. $\eta(x) = -x$ is the repulsion term, and $w(x) = e^{-x^2/h^2}$ is a decaying weight function.

4 Experiments

4.1 Experimental Setup

Implementation details. We train mmPoint using the Adam optimizer. The initial learning rate is set to be 0.0001, and is decayed be 0.5 after every 40 epochs. The point number of the input human template point cloud is 256.

Evaluation metric. Following works in the point cloud generation field, we adopt the L1 version of Chamfer distance as the evaluation metric to quantitatively measure the performance of the proposed method.

Dataset. Dense human point cloud generation from mmWave has not yet been extensively researched due to the challenges involved. One of the main difficulties is obtaining ground truth data for human point clouds. Two possible methods include a whole-body scan and a high-accuracy motion caption system, both of which are complicated and expensive. In

Methods	Average	Scene #1	Scene #2	Scene #3	Scene #4	Scene #5
mmMesh [33]	10.65	9.68	13.21	11.39	10.26	8.73
mmPoint(Ours)	2.92	2.78	3.06	3.25	2.88	2.61

Table 1: Quantitative comparsion on the proposed dataset in terms of per-point L1 Chamfer distance$\times 10^2$ (lower is better).

this work, we propose a novel approach to obtain pseudo-ground truth data for human point clouds from single images using 3D mesh reconstruction techniques. This method is both simple and effective, and it enables us to generate pseudo-ground truth data that can be used for training our model. Specifically, we establish our own dataset by building on the HuPR dataset which includes human 2D images and their corresponding mmWave data. We select 58 out of 276 scenes from the original dataset to ensure diversity and representativeness. To obtain the ground-truth human point cloud, we adopt a two-step preprocessing strategy. In Step 1, we utilize Openpose [4, 23, 32] to extract human poses from 2D images. In Step 2, we use the Expressive Body Capture [17] method to generate human meshes from the 2D images and detected poses. With the generated 3D meshes, we can then sample 3D points on them and finally get dense human point clouds.

The resulting dataset contains both the human point cloud and mmWave data, enabling researchers to train and evaluate their models on this task. Compared to other methods that use expensive human motion capture systems to generate ground-truth data, our proposed approach for generating pseudo-ground-truth human point clouds is cost-effective and effective, allowing for broader access to the dataset and accelerating research in this direction.

4.2 Qualitative Results

In this study, we present a novel deep learning approach, mmPoint, for generating dense and high-quality human point clouds from mmWave signals. To demonstrate the superiority of our method, we compare it with an existing approach called mmMesh [33] which uses the traditional method to generate point clouds from mmWave signals. The qualitative analysis is shown in Fig. 4, which reveals compelling differences in the generated point clouds. Specifically, the human point clouds produced by mmPoint exhibit remarkable density and exhibit a high level of detail, resulting in a more accurate representation of human subjects. In contrast, the point clouds generated by mmMesh are characterized by significant sparsity and noise, leading to a compromised depiction of human forms. These findings unequivocally establish the superior performance of our mmPoint in terms of generating dense and high-quality human point clouds from mmWave signals, highlighting its potential for various applications in computer vision and human sensing.

4.3 Quantitative Results

Table 1 presents a quantitative comparison between our proposed method, mmPoint, and mmMesh in terms of per-point L1 Chamfer distance on our proposed dataset. To establish an appropriate testing protocol, we carefully curated a subset of 5 scenes from our dataset. These selected scenes were designated to serve as the testing set for evaluating the performance and generalization capabilities of our proposed method. The results highlight the superior performance of mmPoint in generating more accurate and detailed point clouds. On

Methods	Average
mmPoint (1-step)	3.83
mmPoint (3-step)	2.92

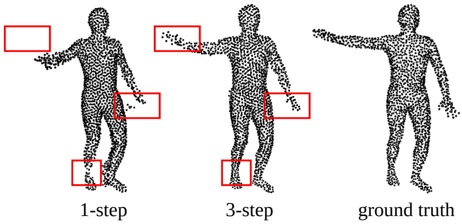

Figure 5: Ablation study on the proposed three-step generation strategy. Both quantitative and visual comparisons are presented.

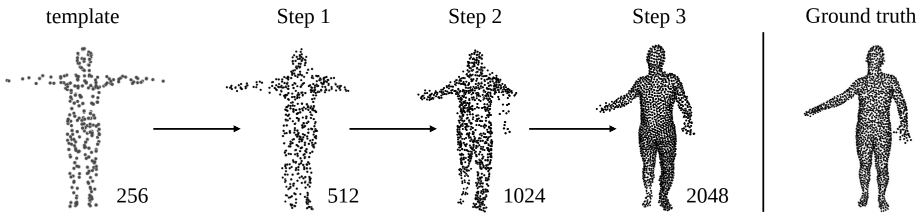

Figure 6: Visualization of generated point clouds via three steps. The number of points is increased by an order of (256,512,1024,2048) through three steps.

average, mmPoint achieves a significantly lower per-point L1 Chamfer distance of 2.92, outperforming mmMesh by a margin of 7.73. This improvement is consistently observed across all four scenes, where mmPoint consistently achieves lower L1 Chamfer distances compared to mmMesh. Notably, in Scene #2, mmPoint achieves a remarkable L1 Chamfer distance of 10.15, indicating its exceptional ability to generate highly precise and dense human point clouds. These quantitative results provide strong evidence of the superior performance of mmPoint over mmMesh, demonstrating its effectiveness in generating high-quality point clouds from mmWave signals.

4.4 Ablation Study

Figure 5 presents the results of our ablation study, where we compare the performance of our mmPoint method using a one-step model (mmPoint (1-step)) with a three-step model (mmPoint (3-step)). For mmPoint (1-step), we use just one LDM and delete the lift operation in it. Moreover, the number of input points is 2048 for mmPoint (1-step). The purpose of this study is to assess the effectiveness of our proposed three-step strategy in generating dense point clouds. The results demonstrate a clear advantage of the three-step model over the one-step model in terms of accuracy and density. The mmPoint (3-step) achieves a significantly lower score, outperforming the mmPoint (1-step) model by 0.91. This improvement in performance highlights the benefits of our three-step strategy, which allows for a more accurate and gradual generation of dense point clouds. To have a deeper understanding of the three-step deformation strategy, we further visualize the generated point clouds of each step in Fig. 6. This figure demonstrates the procedure of point cloud deformation which is consistent with the design intention of the three-step deformation architecture.

5 Conclusion

In this work, we introduced mmPoint, the first model that generates dense human point clouds from mmWave radar signals. Dense human point cloud generation from mmWave is an important and challenging task that has not been extensively researched. We proposed a point cloud deformation approach that simultaneously takes a template human point cloud and mmWave signals as inputs and predicts the corresponding movement for each point to drive the template point cloud to transfer to the target human point cloud. We further established a new dataset consisting of training pairs for the task, which will be made public to accelerate research in this direction. our work presents an important step towards generating dense human point clouds from mmWave radar signals, and we believe that our proposed method and dataset can serve as a valuable resource for future research in this direction.

Limitations. Our proposed approach, for the first time, shows promising results in generating dense human point clouds from mmWave radar signals. However, there are still some limitations that need to be addressed in future research. First, it does not utilize temporal information, which may be important for certain applications involving human motion analysis [40]. Second, the current method is designed to generate a dense point cloud for a single person, which may limit its applicability in scenarios where multiple people are present. Third, although we propose a method to obtain pseudo ground truth for human point clouds from single images, the accuracy of the generated point clouds may be limited by the quality of the input images and the assumptions made during the 3D mesh reconstruction process.

References

[1] Fahad Jibrin Abdu, Yixiong Zhang, Maozhong Fu, Yuhan Li, and Zhenmiao Deng. Application of deep learning on millimeter-wave radar signals: A review. *Sensors*, 21 (6):1951, 2021.

[2] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *International conference on machine learning*, pages 40–49. PMLR, 2018.

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

[4] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017.

[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. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3501–3510, 2022.

[6] Ta-Ying Cheng, Qingyong Hu, Qian Xie, Niki Trigoni, and Andrew Markham. Metasampler: Almost-universal yet task-oriented sampling for point clouds. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, pages 694–710. Springer, 2022.

[7] Jaesung Choe, Byeongin Joung, Francois Rameau, Jaesik Park, and In So Kweon. Deep point cloud reconstruction. *ICLR*, 2022.

[8] L Minh Dang, Kyungbok Min, Hanxiang Wang, Md Jalil Piran, Cheol Hee Lee, and Hyeonjoon Moon. Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, 108:107561, 2020.

[9] Itai Lang, Asaf Manor, and Shai Avidan. Samplenet: Differentiable point cloud sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7578–7588, 2020.

[10] 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. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5715–5724, 2023.

[11] Tao Li, Zhichao Zhao, Yi Luo, Benkun Ruan, Dawei Peng, Lei Cheng, and Chenqi Shi. Gait recognition using spatio-temporal information of 3d point cloud via millimeter wave radar. *Wireless Communications and Mobile Computing*, 2022, 2022.

[12] Chen-Hsuan Lin, Chen Kong, and Simon Lucey. Learning efficient point cloud generation for dense 3d object reconstruction. In *proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[13] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2837–2845, 2021.

[14] Priyanka Mandikal and Venkatesh Babu Radhakrishnan. Dense 3d point cloud reconstruction using a deep pyramid network. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1052–1060. IEEE, 2019.

[15] Luke Melas-Kyriazi, Christian Rupprecht, and Andrea Vedaldi. PC^2: Projection-conditioned point cloud diffusion for single-image 3d reconstruction. *arXiv preprint arXiv:2302.10668*, 2023.

[16] Doğa Nalci and Yusuf Sinan Akgul. Human action recognition with raw millimeter wave radar data. In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pages 1–5. IEEE, 2022.

[17] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[18] Akarsh Prabhakara, Tao Jin, Arnav Das, Gantavya Bhatt, Lilly Kumari, Elahe Soltanaghaei, Jeff Bilmes, Swarun Kumar, and Anthony Rowe. High resolution point clouds from mmwave radar. *arXiv preprint arXiv:2206.09273*, 2022.

[19] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[20] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

[21] Kun Qian, Zhaoyuan He, and Xinyu Zhang. 3d point cloud generation with millimeter-wave radar. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–23, 2020.

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

[23] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*, 2017.

[24] Akash Deep Singh, Sandeep Singh Sandha, Luis Garcia, and Mani Srivastava. Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar. In *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, pages 51–56, 2019.

[25] Yue Sun, Honggang Zhang, Zhuoming Huang, and Benyuan Liu. R2p: A deep learning model from mmwave radar to point cloud. In *Artificial Neural Networks and Machine Learning–ICANN 2022: 31st International Conference on Artificial Neural Networks, Bristol, UK, September 6–9, 2022, Proceedings, Part I*, pages 329–341. Springer, 2022.

[26] Abdelrahman Taha, Qi Qu, Sam Alex, Ping Wang, William L Abbott, and Ahmed Alkhateeb. Millimeter wave mimo-based depth maps for wireless virtual and augmented reality. *IEEE Access*, 9:48341–48363, 2021.

[27] Junshu Tang, Zhijun Gong, Ran Yi, Yuan Xie, and Lizhuang Ma. Lake-net: topology-aware point cloud completion by localizing aligned keypoints. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1726–1735, 2022.

[28] Girish Tiwari and Shalabh Gupta. An mmwave radar based real-time contactless fitness tracker using deep cnns. *IEEE Sensors Journal*, 21(15):17262–17270, 2021.

[29] Shuai Wang, Dongjiang Cao, Ruofeng Liu, Wenchao Jiang, Tianshun Yao, and Chris Xiaoxuan Lu. Human parsing with joint learning for dynamic mmwave radar point cloud. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(1):1–22, 2023.

[30] Yizhou Wang, Zhongyu Jiang, Xiangyu Gao, Jenq-Neng Hwang, Guanbin Xing, and Hui Liu. Rodnet: Radar object detection using cross-modal supervision. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 504–513, January 2021.

[31] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019.

[32] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *CVPR*, 2016.

[33] Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. 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*, pages 269–282, 2021.

[34] Baiju Yan, Peng Wang, Lidong Du, Xianxiang Chen, Zhen Fang, and Yirong Wu. mmgesture: Semi-supervised gesture recognition system using mmwave radar. *Expert Systems with Applications*, 213:119042, 2023.

[35] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. Pu-net: Point cloud upsampling network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2790–2799, 2018.

[36] Tao Yu, Yoshitaka Takaku, Yohei Kaieda, and Kei Sakaguchi. Design and poc implementation of mmwave-based offloading-enabled uav surveillance system. *IEEE Open Journal of Vehicular Technology*, 2:436–447, 2021.

[37] Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. Pointr: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12498–12507, 2021.

[38] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *2018 international conference on 3D vision (3DV)*, pages 728–737. IEEE, 2018.

[39] Peijun Zhao, Chris Xiaoxuan Lu, Jianan Wang, Changhao Chen, Wei Wang, Niki Trigoni, and Andrew Markham. mid: Tracking and identifying people with millimeter wave radar. In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 33–40. IEEE, 2019.

[40] Jia-Xing Zhong, Kaichen Zhou, Qingyong Hu, Bing Wang, Niki Trigoni, and Andrew Markham. No pain, big gain: classify dynamic point cloud sequences with static models by fitting feature-level space-time surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8510–8520, 2022.

[41] Kaichen Zhou, Lanqing Hong, Changhao Chen, Hang Xu, Chaoqiang Ye, Qingyong Hu, and Zhenguo Li. Devnet: Self-supervised monocular depth learning via density volume construction. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*, pages 125–142. Springer, 2022.