Compressed Representation for 3D Human Pose Estimation using WiFi signal

Qijia Zheng[1,2], Bin Liu[1,2(✉)], Jingyang Huang[3], and Nenghai Yu[1,2]

[1] University of Science and Technology of China, Hefei, China
flowice@ustc.edu.cn
[2] Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei, China
[3] School of Computer and Information, Hefei University of Technology, Hefei, China

Abstract. Human pose estimation technology has been increasingly applied in many fields, such as user authentication, activity recognition, and health monitoring. The existing human pose estimation methods are mainly based on cameras and wearable sensors, while some works try to use WiFi signal to perceive fine-grained human joints in recent years. WiFi antennas collects 1D WiFi signal, while final outputs are 3D images containing human joints. However, directly mapping 1D WiFi signal to 3D image with human skeletons has limitations such as degree of accuracy and high costs of computation resources. In this paper, we present a novel approach to bridge the gap between WiFi signal and images for single person 3D human pose estimation using commercial WiFi devices, named CRPose. We propose to use high resolution heatmap to model joint locations, devising an effective compressed method to extract the posture information embedded in 1D WiFi signal. CRPose takes the 1D WiFi data as input and supervised by a teacher network, which outputs the dense representations of 3D human postures. Our experimental evaluation shows that our method on a real-world WiFi sensing testbed with distributed antennas performs favorably when compared to state of the art methods on WiFi-based 3D human pose estimation. CRPose can localize each joint on the human skeleton with an average error of 5.1cm, achieving a 55% improvement in accuracy and a 64% time saving in inference stage over the state-of-the-art posture construction model designed for radio frequency radar sensors.

Keywords: Wireless sensing · Human pose estimation · Channel state information

1 Introduction

In recent years, Human Pose Estimation (HPE) has seen significant progress, as well as wireless sensing system, mainly thanks to machine learning technique. HPE is considered promising in many scenarios such as patient activity monitoring in hospital and suspicious behavior detection in public places. Camera based HPE methods are more pervasive due to intuitionistic and abundant information of images, while posing a significant risk of leaking sensitive information to users, thereby bringing a threat to users' privacy. The methods based on radar sensors and WiFi sensors are able to monitor users without collecting visible information of human body, thus protecting their privacy. However, it has been found that radar based solutions are too expensive to be truly deployed in daily life.

In the meantime, various WiFi sensing systems and algorithms have been proposed to track the position of the monitored human subject and recognize people's activities through analyzing the signals reflected off and penetrated the human body. With the localization and recognition accuracy progressively increased, some pioneer study offers fine-grained human body sensing system solution[2,3], demonstrating that with the supervision of visual information, radio frequency (RF) signals can be used to generate 2D and even 3D skeletal representations of the human body. By overcoming the technical challenges faced by traditional camera based and dedicated radar sensors based human perception solutions, such as occlusion, poor lighting, privacy issues, as well as cost, WiFi signal based human pose estimation technique demonstrates the potential to enable a new generation of applications capable of supporting more sophisticated human monitoring and interactions. However, mapping 1D WiFi Signal to 2D/3D images in an end-to-end manner is an ill-posed question[9]. Thereby, a fundamental question rise: How to efficiently extract the human activity information in WiFi signal? To answer this question, we propose to make use of the pervasive WiFi devices, and "image" 3D human postures from WiFi signals. More specifically, we aim to predict 3D human pose composed of 14 joints (i.e., head, left ankle and right shoulder) and then associate them into 3D skeleton (i.e. arms and legs) of human body. And, we present a novel strategy to represent WiFi data and train our neural network. We empirically follow the intuition that remapping the ground truth to an compressed intermediate representation is able to efficiently extract the spatial information from sparse WiFi signal. For this purpose, self-supervised networks such as autoencoders represent a natural choice for searching for intermediate representation. The core of our proposal relies on the creation of an alternative ground-truth representation that preserves the most informative content of the original ground-truth. Specifically, our WiFi-based HPE pipeline consists of two modules: At first, the pre-trained teacher network is leveraged to obtain a denser representation of the WiFi signals. This new compressed representation is used as the target ground-truth during our network training. Then, at inference stage, input WiFi signal is fed into model and predict the keypoints of human body. To summarize, the main contributions of our proposal are:

– We build a sensing system using commercial WiFi devices and a web camera to simultaneously collect Channel State Information(CSI) and images with the synchronization error less than 1ms.
– A two stages deep learning framework based on autoencoders framework is proposed to map WiFi signal to human joint coordinates.

- We propose a simple and effective method that maps sparse CSI to more denser representation, which saves computational resources while extracting most of the informative content from WiFi signal.

The rest of this paper is organized as follows: First, we take a review of related works in section 2. Next, we give an overview of the system design and describe the proposed deep learning framework in sections 3. Then, evaluation of the system is presented in Section 4. Finally, we conclude our work in Section 5.

2 Related Work

Human pose estimation refers to the detection of important body joints or key points from images or videos. At the same time, associating the detected key points to form the joints of human body is also an important step after detecting these key points. In this section, we will briefly introduce some popular HPE methods in recent years, which can be divided into camera-based, wearable sensor-based and RF-based methods.

Camera-based. Considering the number of people in the image or video frame, single person pose estimation and multi person pose estimation can be distinguished [8]. Single person pose estimation only involves predicting the pose of a single person in the entire image or video. The multi person pose estimation task is more challenging, as the position of the arms or limbs in the image is unknown. To solve this problem, the most commonly used methods are: (1)top-down method, which is the easiest method for human detection; Predict each person's body parts and calculate posture; And (2)a bottom-up approach, including detecting all human parts in images or videos, and then grouping all body parts belonging to a specific person. Most recent work has focused on the design of posture decoders, with increasing emphasis on exploring contextual information and inherent features of body structure. Toshev et al. [16] proposed DeepPose, which is one of the first human pose estimation methods based on deep convolutional neural networks (DCNN). Through a series of pose predictors based on DCNN, DeepPose formalizes the task of estimating human key point estimation into a regression problem.

Wearable Sensor-based. The VICON (Vicon Motion Systems) system actively estimates human body status by wearing optical markers and inertial sensors on users, while Optitrace uses passive optical markers. These methods provide accurate results. The main drawback of these systems is the need for a structured environment. As another example, exoskeleton sensors can accurately estimate human posture due to their rigid structure and high-quality sensor system, but the accompanying drawbacks are high cost and high equipment weight. Compared to camera based pose estimation methods, these methods perform better in scenarios such as poor lighting conditions, but they typically have high costs and are not convenient for daily wear, nor can they be used in unstructured or outdoor scenes. The above wearable sensor based solution provides very accurate pose and motion estimation methods, which can perform well in unstructured and outdoor environments.

RF-based. Starting from traditional behavior recognition tasks and respiratory monitoring tasks[1,4,7,13], the accuracy of RF perception has gradually improved. Zhao et al. [2,3] used Frequency Modulated Continuous Wave (FMCW) radar for the first time to reconstruct 2D and 3D human skeleton, as well as demonstrating the ability of this scheme to perceive fine-grained human pose through wall. Li et al. [10] used FMCW radar to perform human behavior recognition under extreme light (darkness) and occlusion conditions. Although the radar signal showed amazing results, there is still no publicly available dataset in the field of wireless signal based attitude estimation, which has led to slow progress in research on fine-grained perception of wireless signals. To overcome this problem, Cai et al. [15] proposed a system for human motion perception using RF signals. This system utilizes recognized human behavior actions to generate corresponding RF signals in reverse, thereby supplement the training data for RF-based scheme. For the first time, Wang [9]and Jiang[5]separately design a 2D and 3D human pose estimation system using commercial WiFi devices, demonstrating the feasibility of 2D and 3D human skeleton reconstruction using WiFi signal without dedicated RF sensors. Afterwards, Wang and Ren et al.[6,11,14] used signal processing methods to design a 3D human pose estimation system based on WiFi signals.

3 Method

The following subsections summarize the key components of CRPose. Section 3.1 provides a brief overview of the entire WiFi sensing system. Next, in Section 3.2, we describe signal preprocessing methods employed. Finally, Section 3.3 illustrates our proposed neural network which transit the low-dimensional WiFi signal to the high-dimensional image by generating a compact and more tractable representation.

3.1 System Overview

In this paper, we aim to accurately estimate the subject's 3D pose using the WiFi signals that carry human motion information collected through wireless devices. As shown in Figure 1, the WiFi signal 3D pose estimation system mainly comprises three modules: the Data synchronization collection module, the Signal Pre-processing module, and the Pose Estimation module. Firstly, To capture both human body images in indoor and outdoor environments and WiFi data affected by human body movements, we construct a dual-mode data collection system comprising multiple WiFi signal receivers, WiFi transmitters, and a single RGB camera. Next, the Signal Pre-processing module performs several steps to enhance the quality of the collected CSI data. It filters the amplitude of CSI to eliminate small-scale changes in the WiFi signal that may arise from environmental interference. Additionally, the module eliminates any phase offset present in the data. Finally, the pose estimation module proposes a two-stage deep neural network based on an encoder decoder structure, which extracts temporal and spatial information from CSI through the pose encoder. The output pose features obtain pseudo 3D coordinates of key points through the pose decoder; Finally, combined with the RGB camera internal parameters in the local scene, the 3D skeleton map of the human body was obtained by reverse projection to the real world 3D coordinates.

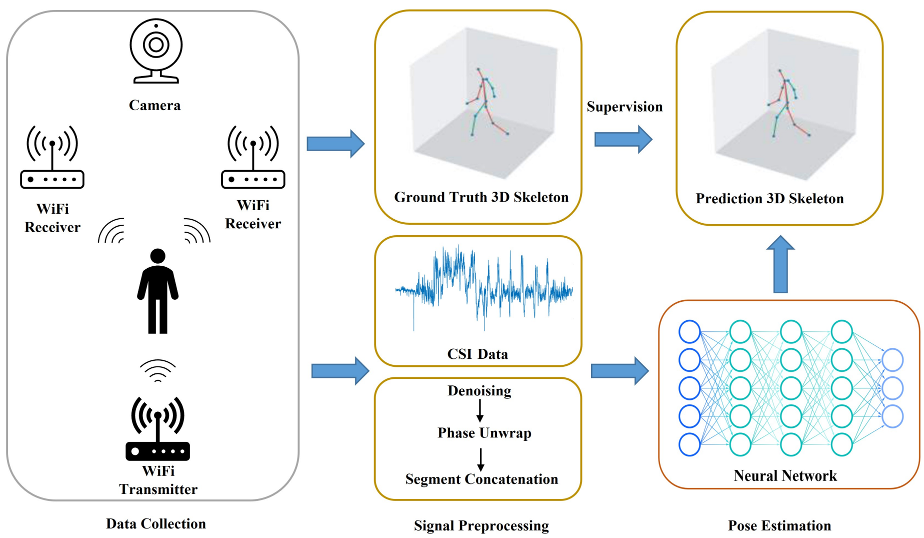

Fig. 1. System Overview

3.2 Signal Preprocessing

WiFi receivers continuously measure CSI in WiFi NICs, and feed back to transmitters. CSI characterizes the frequency response of wireless channel(CFR), and can be described as formula.

$$H(f;t) = \sum_{n}^{N} a_n(t) e^{-j2\pi f \tau_n(t)}, \tag{1}$$

where $a_n(t)$ is the complex valued representation of attenuation ,$\tau_n(t)$ is ,and f is the carriers frequency.

The CSI stream measurement provided by commercial wireless network cards is extremely noisy. The source of noise in CSI streams is the internal state transition between the wireless WiFi network cards of the transmitter and receiver, such as transmission power adjustment caused by transmission power control algorithms, transmission rate adaption caused by transmission rate adaptive algorithms, and changes in CSI internal reference levels. These parameters are rooted in the MAC802.11 subsystem and serve WiFi communication, with considerable randomness. This type of internal state transition generates high amplitude pulses and burst noise in CSI. In addition to the interference caused by the aforementioned hardware and software systems on WiFi signals, there are also various types of noise mixed with WiFi signals when they propagate in space. For example, multipath effects, shadow fading, and co-channel interference from other WiFi devices. Eventually, we adopt the Butterworth filter performing low-pass filtering on the original CSI signal to remove high-frequency noise caused by such situations.

$$H(s) = \frac{1}{1 + \left(\frac{s}{\omega_c}\right)^{2n^{0.5}}} \tag{2}$$

where s is the complex frequency variable. ω_c is the cutoff frequency, which defines the point at which the filter attenuates the signal by a certain amount. n is the order of the filter, determining the sharpness of the roll-off in the stopband.

For phase processing, the raw CSI measurement consist of complex elements $z = a + bi$. The phase $\varPhi$ are calculated using formula $\varPhi = \arctan(b/a)$, and the phase will be wrapped when the phase value is not included in $-\pi$ to π. So, we clean the phase by unwrapping the phase using formula below:

$$\begin{aligned} \varDelta\phi_{i,j} &= \varPhi_{i,j+1} - \varPhi_{i,j} \\ \text{if } \varDelta\phi_{i,j} > \pi, \varPhi_{i,j+1} &= \varPhi_{i,j} + \varDelta\phi_{i,j} - 2\pi \\ \text{if } \varDelta\phi_{i,j} < -\pi, \varPhi_{i,j+1} &= \varPhi_{i,j} + \varDelta\phi_{i,j} + 2\pi \end{aligned} \tag{3}$$

3.3 Neural Network

The proposed CRPose deep learning framework is illustrated in Figure 1. After signal preprocessing, we transformed the raw CSI data extracted from M distributed antennas into a sequence of input data, and feed into the CRPose model to predict 3D human joints. The following subsections present the details of proposed deep learning framework.

(1) CSI transition module. To bridge the information representation gap between one-dimensional WiFi data and three-dimensional heatmap data, we introduce a CSI transition module, as well as the pose encoder, that effectively maps the input from the CSI domain to the image domain. The input of the CSI transition module is represented by 5 CSI frames, C, while its output $f(C)$, aims to predict the intermediate representation obtained with the $e(I)$. CSI contains both temporal and spatial information, . The specularity of WiFi signals[9] may lead to the loss of human motion information at a specific frame, resulting in the loss of human key point information in the heatmap. Therefore, CSI transition module adopted the Transformer[19] structure to extract the temporal information from CSI. Transformer which has achieved excellent performance in the

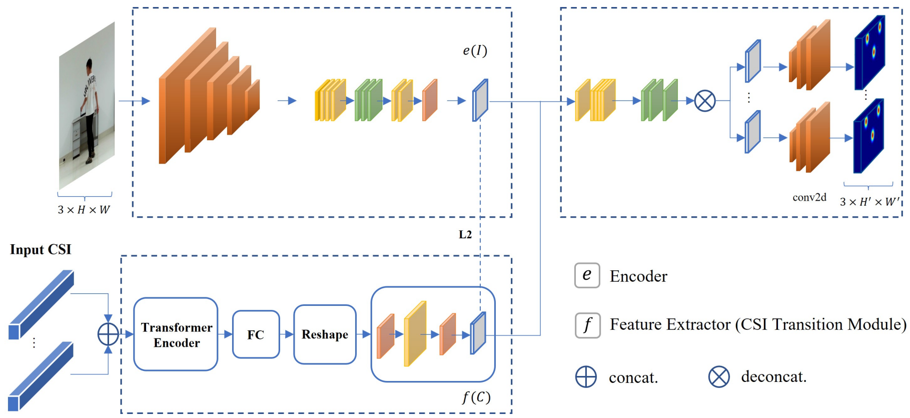

Fig. 2. Schematization of the proposed model pipeline. At training time, the Encoder e produces the intermediate $e(I)$ which are used as ground truth from the CSI feature extractor f. At test time, the intermediate representation $f(I)$ computed by the CSI feature extractor is fed to the Decoder d for the keypoints heatmap output.

translation quality of natural language processing(NLP). The Transformer module followed the encoder-decoder structure using stacked self-attention and pointwise, fully connected layers. The model multiplied the input vector with three different weight matrices to obtain the queries vector(Q), keys vector (K) and values vector(V). After that, we employ convolutional neural networks(CNNs) to extract spatial features. In particular, we use one layer of transposed convolution and two convolutions followed by rectified linear functions(ReLU) to add non-linearity to the model.

(2) Training strategy. Inspired by LoCO[17], our method use volumetric heatmaps to represent human keypoints locations. Gaussian

$$\mathbf{H_j}(\mu) = e^{-\frac{\|\mu-\mu_j\|^2}{\sigma^2}} \tag{4}$$

We leverage the encoder-decoder structure based model, which is pre-trained on the virtual multi-person dataset, to get the ground truth volumetric heatmaps of human keypoints. This type of ground truth representation called volumetric heatmaps is in a compressed data representation that can efficiently extract the information and avoid the disadvantages embedded in sparse WiFi signal. The goal of our proposed training strategy is therefore to learn a compressed representation of the input volumetric heatmaps that preserve human pose information content, which results in the preservation of the position of the various joints in the original maps. We trained the CSI transition module by minimizing the MSE loss between$f(C)$ and $e(I)$, where I is the image associated with the CSI, C.

(3) Pose decoder. The Pose Decoder aims to decode the compressed volumetric heatmaps to final keypoints heatmap output. At inference time, the pseudo-3D coordinates of human joints are acquired from Pose Decoder through a local maxima search. With the intrinsic camera parameters, the true coordinates of the joints can be recovered. Then, Pose Decoder aims to decode the compressed representation of CSI transition module to keypoint heatmaps. We use 14 joints representation for . Furthermore, human pose estimation method based on cameras may encounter prediction failure cases and joint positioning errors when predicting the final joint points due to the probability of lighting, occlusion. Pose estimation models based on WiFi also face similar problems. During the transmission process of WiFi signals, the superposition of WiFi signals caused by multipath effects at the receivers blurs the CSI measurement, which may result in missing joints cases. We further adapt a MLP network to refine the predicted 3D poses.

4 Experiments

4.1 Testbed

Data Collection and Annotations. Our WiFi testbed consists of one WiFi router and three laptops. These devices are divided into two group, one for IEEE802.11n WiFi data collection, another for IEEE802.11ax WiFi data collection. Two laptops are equipped with Intel 5300 wireless NIC connected with three antennas. One laptops are equipped with Intel AX200 wireless NIC connected with two antennas. We use the WiFi router as the transmitter and laptops as receivers. On our testbed, Linux 802.11n CSI tools[20] are used to log IEEE802.11n CSI data and PicoSence to log IEEE802.11ax CSI data. We enable the transmitter to send ICMP packets to the receivers. Thus, all the receivers can simultaneously receive packets from the transmitter. WiFi signals are set on channel 5 GHz where there is little interference from other devices. The packet rate is set at 100 packets per second. We use a web camera connected to Ubuntu20.04 to capture the human pose images and generate ground truth human skeleton keypoints. XDP(express data path)[18]is used to redirect the ICMP packets to Linux user space and record the timestamp of packets. PTP is used to synchronize the system time among laptops, the average synchronization error is less than 1ms.

4.2 Dataset

In our experiments, We collect 345000 samples of 3D skeleton frames and 1725000 WiFi CSI packets correspondingly in 2 different environments. To evaluate the system performance, subjects are asked to walk around without any specific instruction.

4.3 Evaluation Results

Qualitative results. Table 1 shows the average positioning error for each joint point, and also calculates the average error for all joint points. From the table,

Fig. 3. Testbeds and the basic scenario of WiFi based pose estimation

it can be seen that the overall positioning error of the proposed method in this article is approximately 5.1cm, while the overall positioning error of the baseline method RFPose is 9.3cm. Moreover, for some specific joint points, like RH (Right Hand), RA (Right Ankle), RK (Right Knee), LK (Left Knee), LA (Left Ankle), LH (Left Hand) have relatively larger localization error values, while joint points located in the middle of the body, such as H (Head), N (Neck), RS (Right Shoulder), LS (Left Shoulder), have relatively less localization error values. According to a simple analysis, the reason for this situation is that the middle part of the human body is located in the center of the Fresnel region formed by the WiFi signal transmitter and receiver [1,4]. In our experimental environment, the Fresnel region formed is an ellipsoid with a long half axis of about 5m and a short half axis of about 0.4m. The middle part of the human body is located on the direct path with the strongest signal power and the least interference, while the leg and foot joints are located at relative edge positions. Due to multipath effects, the limb morphology of the leg and foot joint positions is relatively complex, making the propagation path of WiFi signals in this part more complex. In order to showcase the generalization performance of our models across various WiFi protocols, we present the prediction results obtained from two commonly used WiFi signals in the table2. The results indicate that the IEEE802.11ax WiFi signal, characterized by an increased number of subcarriers and finer signal granularity, exhibits comparatively higher performance.

Table 1. The average keypoint location error of 14 joints (unit:cm)

Joints	H	N	RS	RE	RW	LS	LE	LW	RH	RK	RA	LH	LK	LA	Overall
RFPose	6.6	6.4	8.4	10.1	11.4	8.9	9.9	11.1	8.2	6.4	8.9	10.9	9.9	12.3	9.3
Ours	4.2	4.1	3.8	5.7	4.3	4.1	6.4	5.4	2.1	6.2	7.6	3.7	6.3	8.0	5.1

Quantitative results. Figure 4 shows the quantitative results. The first line in the figure shows the raw image data collected by the RGB camera. The second row displays the corresponding GT data predicted by the monocular

Table 2. The average keypoint location error under the IEEE802.11n and IEEE802.11ax protocol (unit:cm)

Joints	H	N	RS	RE	RW	LS	LE	LW	RH	RK	RA	LH	LK	LA	Overall
IEEE802.11n+RFPose	7.6	7.0	8.1	10.0	11.1	9.1	9.2	10.9	8.4	6.6	9.2	10.3	9.9	11.8	9.9
IEEE802.11ax+RFPose	6.6	7.4	7.8	9.5	10.6	9.8	9.0	11.5	8.9	5.9	8.9	10.0	10.2	11.3	9.1
IEEE802.11n+Ours	4.9	4.5	4.8	5.6	3.9	4.7	5.9	5.8	3.1	5.8	6.0	4.3	6.1	7.0	5.6
IEEE802.11ax+Ours	3.9	3.9	4.1	5.2	3.8	4.7	6.0	5.1	2.3	6.0	6.3	3.9	6.5	7.3	4.9

pose estimation model, and the third row displays the attitude prediction results obtained by the model proposed in this article based on CSI. At the pictures, it can be seen that the model performs well under different human posture.

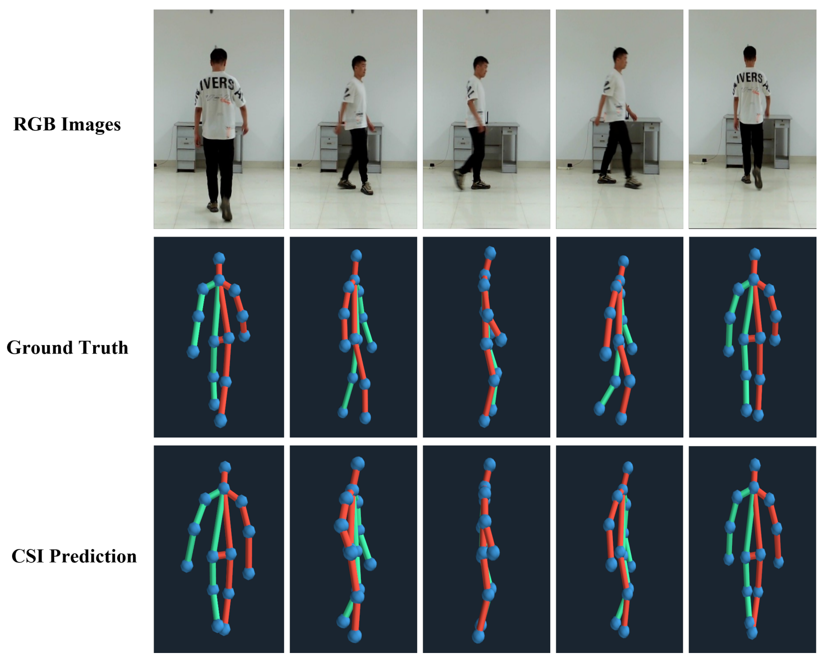

Fig. 4. Qualitative results of our proposed method. The first row shows the RGB image captured by the web camera, the second row shows the prediction Ground Truth using the pre-trained model, and the third row shows the our model's prediction based on WiFi signals.

Runtime Analysis. Compared to 2D and 3D image data, 1D WiFi data has a smaller data scale. We provide a comparison of runtime between WiFi based pose estimation models and 2D image based pose estimation models. Our model take less convolution operations which brings less runtime cost.

Table 3. Runtime of inference stage of 1 second WiFi data on a single NVIDIA Titan XP GPU (unit:s)

Model	Pose Encoder	Pose Decoder	Overall
RFPose	-	-	1.63
ours	0.13	0.92	1.05

5 Conclusion

In this paper, we designed a system for 3D human pose estimation based on WiFi signals, which uses commercial WiFi devices to construct fine-grained human pose images. A new representation of wireless data was proposed, and different training methods were used to train our WiFi-based human pose estimation model. Our approach allows us to exploit a denser representation of WiFi signal as a ground truth for the WiFi-based 3D Human Pose Estimation task. Our method consists of two stages. One is to train an Pose Encoder based on compressed pose representation, which is mainly used to encode WiFi data to an intermediate representation for Pose Decoder. Another is to decode the compressed WiFi information to human joints at inference stage. The experimental results show that the human pose images constructed by our system strictly match real image data, and have lower computational resource consumption compared to 2D and 3D convolution operations on images, which proves the effectiveness our method.

References

1. Niu, K., Zhang, F., Chang, Z., Zhang, D.:From Fresnel Diffraction Model to Fine-grained Human Respiration Sensing with Commodity Wi-Fi Devices. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, pp. 416-419 (2018)
2. Zhao, M., Tian, Y., Zhao, H., Alsheikh, M. A., Li, T., Hristov, R., et al.: RF-based 3D skeletons. In: Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, pp. 267-281. (2018)
3. Zhao, M., Li, T., Abu Alsheikh, M., Tian, Y., Zhao, H., Torralba, A., Katabi, D.: Through-Wall Human Pose Estimation Using Radio Signals. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7356-7365. (2018)
4. Zeng, Y., Wu, D., Xiong, J., Yi, E., Gao, R., Zhang, D.: FarSense: Pushing the Range Limit of WiFi-based Respiration Sensing with CSI Ratio of Two Antennas. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, pp. 1-26. (2019)
5. Jiang, W., Xue, H., Miao, C., Wang, S., Lin, S., Tian, C., Murali, S., Hu, H., Sun, Z., Su, L.: Towards 3D human pose construction using WiFi. In: Proceedings of the 26th Annual International Conference on Mobile Computing and Networking, pp. 1-14. (2020)

6. Wang, Y., Guo, L., Lu, Z., Wen, X., Zhou, S., Meng, W.: From Point to Space: 3D Moving Human Pose Estimation Using Commodity WiFi. In: IEEE Communications Letters, pp. 2235-2239. (2021)
7. Wang, W., Liu, A. X., Shahzad, M., Ling, K., Lu, S.: Understanding and Modeling of WiFi Signal Based Human Activity Recognition. In: Proceedings of the 21st annual international conference on mobile computing and networking, pp. 65-76. (2015)
8. Munea, T. L., Jembre, Y. Z., Weldegebriel, H. T., Chen, L., Huang, C., Yang, C.: The progress of human pose estimation: a survey and taxonomy of models applied in 2D human pose estimation. In: IEEE Access, pp. 133330-133348. (2020)
9. Wang, F., Zhou, S., Panev, S., Han, J., Huang, D.: Person-in-WiFi: Fine-Grained Person Perception Using WiFi. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5452-5461. (2019)
10. Li, T., Fan, L., Zhao, M., Liu, Y., Katabi, D.: Making the Invisible Visible: Action Recognition Through Walls and Occlusions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 872-881. (2019)
11. Ren, Y., Wang, Z., Wang, Y., Tan, S., Chen, Y., Yang, J: GoPose: 3D Human Pose Estimation Using WiFi. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, pp. 1-25. (2022)
12. Qian, K., Wu, C., Zhou, Z., Zheng, Y., Yang, Z., Liu, Y.: Inferring Motion Direction using Commodity Wi-Fi for Interactive Exergames. In: Proceedings of the 2017 CHI conference on human factors in computing systems, pp. 1961-1972. (2017)
13. Qian, K., Wu, C., Zhang, Y., Zhang, G., Yang, Z., Liu, Y.: Widar2.0: Passive Human Tracking with a Single Wi-Fi Link. In: Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services, pp. 350-361. (2018)
14. Yu, C., Wu, Z., Zhang, D., Lu, Z., Hu, Y., Chen, Y.: RFGAN: RF-Based Human Synthesis. In: IEEE Transactions on Multimedia. (2022)
15. Cai, H., Korany, B., Karanam, C. R., Mostofi, Y.: Teaching RF to Sense without RF Training Measurements. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, pp. 1-22. (2020)
16. Toshev, A., Szegedy, C. Deeppose: Human pose estimation via deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp. 1653-1660. (2014)
17. Fabbri, M., Lanzi, F., Calderara, S., Alletto, S., Cucchiara, R.:Compressed Volumetric Heatmaps for Multi-Person 3D Pose Estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7204-7213. (2020)
18. Høiland-Jørgensen, T., Brouer, J. D., Borkmann, D., Fastabend, J., Herbert, T., Ahern, D., Miller, D.: The eXpress Data Path: Fast Programmable Packet Processing in the Operating System Kernel. In: Proceedings of the 14th international conference on emerging networking experiments and technologies, pp. 54-66. (2018)
19. Waswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, et al.: Attention is all you need. In: NIPS. (2017)
20. Halperin, D., Hu, W., Sheth, A., Wetherall, D.: Tool release: Gathering 802.11 n traces with channel state information. In: ACM SIGCOMM computer communication review, pp. 53-53. (2011)