2022 IEEE 12th International Conference on RFID Technology and Applications (RFID-TA) | 978-1-6654-6594-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/RFID-TA54958.2022.9924133

RFPose-GAN: Data Augmentation for RFID based 3D Human Pose Tracking

Chao Yang, Ziqi Wang, and Shiwen Mao
Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA
Email: {czy0017, zzw0104}@auburn.edu, smao@ieee.org

***Abstract*—In the age of Artificial Intelligence of Things (AIoT), human pose tracking has attracted increasing interest in many fields. To address the limitations of conventional vision based pose tracking techniques, Radio Frequency (RF) based pose monitoring has been proposed in recent years. However, most of the existing RF-based approaches depend on a vision-aided multi-model learning model, which requires extensive labeled data for supervised training. Collecting such large amounts of training data is time-consuming and costly. In this paper, we address this issue by proposing a Generative Adversarial Network (GAN) based data augmentation method, termed RFPose-GAN, to generate synthesized datasets to assist the training of multi-model neural networks. Our experimental results demonstrate the efficacy of the proposed data augmentation approach on improving the performance of 3D human pose tracking when there is only a limited amount of training data.**

***Index Terms*—Radio-frequency identification (RFID), 3D human pose tracking, Generative Adversarial Network (GAN), Data augmentation.**

I. Introduction

With the fast development of Artificial Intelligence of Things (AIoT), human pose tracking has been widely used in many application domains, such as safety surveillance, human-computer interaction, and somatosensory gaming [1]. Computer vision-based techniques have been shown effective in many cases, but their performance is usually constrained by the illumination condition, camera angle, and obstacles in the line-of-sight path. Using a video camera frequently raises privacy concerns. To address these limitations, researchers have resorted to radio frequency (RF) sensing based solutions, where various RF technologies have been exploited, such as WiFi, Frequency Modulated Continuous Wave (FMCW) radar [2], and Radio Frequency Identification (RFID) [3]. Although effectively addressed all the above issues, RF sensing-based solutions also bring about many new challenges, such as noisy and sparse RF data, susceptible to environmental interference, and difficulty in extracting human motion related features from RF data. It has been shown that a multi-model learning approach would be useful, where vision data is used for supervised training of the model to learn human activity features from RF input data [2], [3].

To be resilient to environmental interference, the near-field communication technology, RFID, has been utilized for human pose tracking, where RFID tags are used as low-cost wearable sensors [4]. The several existing RFID-based pose tracking systems have demonstrated the feasibility and high potential of this approach. In such systems, passive tags are attached to the human body, so that the collected RFID phase data can carry the features of body movements. For example, the RFID-Pose system [4] adopts the vision-aided multi-model learning design, where a deep kinematic neural network is trained with the synchronized data sequences sampled simultaneously by a Kinect 2.0 device and an RFID reader, respectively. After training the deep learning model, RFID-Pose can learn body motion related features from RFID data and reconstruct a 3D human pose in real time without needing vision data anymore.

One downside of the vision-aided method is that the training process requires a large amount of synchronized, or paired, RFID and Kinect data. The dataset collection is time-consuming; the test subject needs to keep on performing many activities for an extended period of time in front of the Kinect camera and RFID reader. In addition, to improve the generalizability to different deployment environments or different test subjects, the training dataset should possess high diversity with respect to various environments and body forms [5], [6]. One solution to this problem is meta-learning [5], where a model is pre-trained first using a variety of data, and then fine-tuned using a small amount of new data when deployed in a new environment. Alternatively, we propose an effective data augmentation approach to generate a large amount of synthesized training data to avoid the cost of dataset collection [7], [8]. We find that generating human pose data may not be that expensive, because pose data can either be extracted from videos or simulated by a 3D animation software [9]. To this end, the key challenge is how to generate the synthesized RFID phase data that is paired with a given human pose sample.

In this paper, we address the above challenge by developing a Generative Adversarial Network (GAN) based data augmentation system, termed RFPose-GAN, to generate synthesized training data at a low cost. The RFPose-GAN system requires a given pose dataset. It will then synthesize RFID data samples paired with each sample in the pose dataset. As discussed, the pose data can be sampled using a camera or even be simulated, so collecting the human pose dataset will be inexpensive. Then, we use RFPose-GAN to generate synthesized RFID phase data sequence, which is synchronized with the given pose data sequence. Such synthesized, paired data generated by RFPose-GAN will be useful for supervised training of the multi-model deep learning network used in RFID-based human pose tracking systems, to achieve a good performance when only a small amount of real sampled data is available. Through an experimental study, we demonstrate that

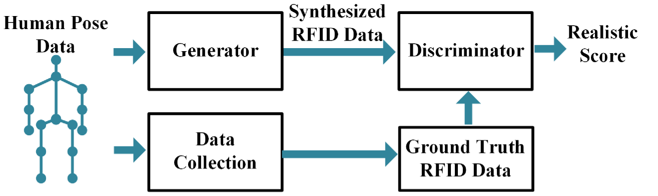

Fig. 1. Overview of the proposed RFPose-GAN system.

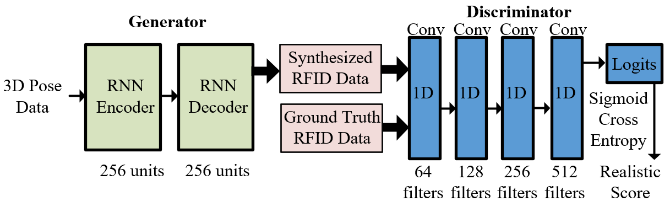

Fig. 2. The generative adversarial network model designed for RFPose-GAN.

the proposed RFpose-GAN can effectively reduce the amount of real sampled data required for model training, so the cost of training dataset collection will be significantly reduced.

The main contributions of this paper are summarized as:

- To the best of our knowledge, RFPose-GAN is the first data augmentation system to effectively reduce the cost of training data collection for RFID-based pose tracking.
- We design the GAN-based model to generate synchronized RFID data paired with the given human pose data. The generator in the GAN model is designed to extract features from the human pose data, while the discriminator is responsible for identifying whether the generated fake data is similar to the real synchronized RFID data.
- The proposed data augmentation system is implemented and evaluated with extensive experiments. The results show that the proposed data augmentation technique can effectively enhance the performance of 3D human pose tracking with only a small amount of real sampled data.

The rest of this paper is organized as follows. We discuss the RFPose-GAN design in Section II. In Section III, we present our experimental study. Section IV concludes this paper.

II. Design of RFPose-GAN

A. System Overview

The architecture of RFPose-GAN is presented in Fig. 1. As the figure shows, the main idea is to train an optimized generator, which can generate synthesized RFID data from the given human pose data. For offline training of the GAN model, we first collect synchronized RFID phase data and human pose data. In the generator, a recurrent autoencoder is used to learn the body movement features from the human pose data. Then the extracted feature is translated into fake RFID data. However, the generated fake RFID phase data may not be useful for training the deep learning network in RFID-based pose tracking systems. The GAN model has a discriminator to identify whether the generated fake data is similar to the real RFID data. Specifically, the discriminator will calculate a realistic score to identify the similarity between real RFID data and generated fake data. When the realistic score reaches a predefined threshold, the generated fake data will be considered as useful data, i.e., the discriminator cannot decide whether the data is fake or not. During online testing, the well-trained generator will generate synthesized RFID data paired with the given human pose data.

B. Deep Network Design in RFPose-GAN

The detailed architecture of the proposed GAN model is plotted in Fig. 2. As the figure shows, the GAN architecture is leveraged to generate synthesized RFID data [10]. The goal of training the GAN model is to learn and map the features of ground truth RFID data, paired with Kinect captured poses, to synthesized RFID data. The 3D pose data is obtained using Kinect in the form of consecutive 3D coordinates of 12 human joints. The GAN network comprises two key parts: (i) a Recurrent Autoencoder as the generator and (ii) a 1D convolutional neural network as the discriminator. In the Recurrent Autoencoder, the Recurrent Neural Network (RNN) is used for learning and extracting the time-series features of Kinect data, which has a simple yet powerful model structure. In the generator, the features of Kinect data are first extracted by the Recurrent Encoder and kept in the hidden layers, where 256 gated recurrent units (GRU) are incorporated to compute the encoding outputs. The encoding outputs in the previous time slot are then fed into the next encoder along the timeline. As a result, the Recurrent encoder can learn features of Kinect data from both the present time and the previous time. The next step is to leverage the Recurrent decoder to transform the extracted features kept in the hidden layer into synthesized RFID data.

The synthesized RFID data is further evaluated by the discriminator. The proposed discriminator consists of four hidden layers, each being a 1D convolutional layer. 1D convolutional layers are selected because of their ability for extracting temporal features. The other input to the discriminator is the ground truth RFID data that is, in essence, the phase variations collected from 12 RFID tags in consecutive time slots. The first hidden layer has 64 kernels and a leaky ReLU function to compute the layer outputs. The second hidden layer has 128 kernels, the third layer has 256 kernels, and the fourth layer has 512 kernels. Each of the three hidden layers' convolution outputs is fed through a batch normalization function and then a leaky ReLU function. The kernel width of all the layers is set to 1. In the end, the output of the fourth hidden layer is fed into a final 1D convolutional layer to flatten it as a logits vector for realistic score calculation.

Fig. 3 illustrates an example of the synthesized RFID data by the GAN generator. The RFID data here refers to the RFID phase variations in the responses received from each of the 12 tags that are attached to the 12 joints of the test subject. Before this work, RFID phase variations have already been collected and orgznied in the form of 3rd-order tensors. Fig. 3

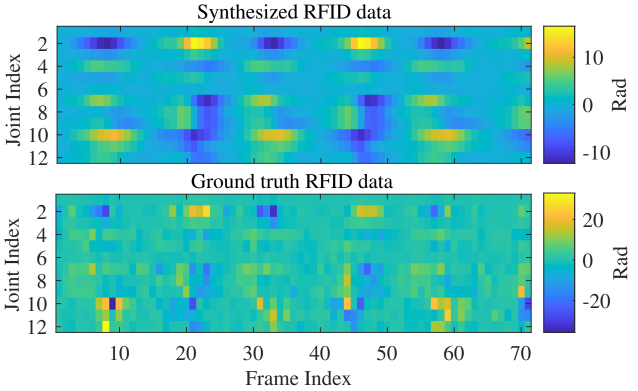

Fig. 3. Example of synthesized RFID data generated by RFPose-GAN.

depicts the phase variations of the 12 joints that are received by one out of the three RFID antennas in 71 consecutive time slots. The upper plot presents the phase variations of the synthesized RFID data, while the lower plot illustrates the phase variations of the ground truth RFID data. We can see that there is considerable overall similarity between these two for each of the joints and over time. This example visually validates that the RFpose GAN model is capable of generating RFID data that can be as good as the ground truth RFID data, as the generator can successfully fool the discriminator with the synthesized RFID data after being well trained.

III. Experimental Evaluation

A. System Configuration and Data Collection

To evaluate the performance of RFPose-GAN, we develop a prototype system with an off-the-shelf Impinj R420 reader, which is used to interrogate passive ALN-9634 (HIGG-3) tags using three S9028PCR polarized antennas. The human pose data used for training the proposed generator is collected with a Kinect 2.0 device. The original dataset is collected for three subjects in five different environments, which will be used to train RFPose-GAN. The augmented dataset is comprised of the original RFID data and the synthesized RFID data generated by RFPose-GAN. Five different types of activities are included in the this study, including boxing, drinking, squatting, hand waving, and walking. It takes about 8 hours to train the RFPose-GAN with 30 training datasets consisting of different movement types with a GTX 1660 Ti Graphics card.

Three datasets consisting of different training data samples are used to highlight the efficacy of data augmentation. The first dataset, *Dataset_3Act*, includes three different types of activities, and the second dataset, *Dataset_5Act*, consists of five different types of activities. Each of these two datasets only include a small amount of samples for each type of activity, which are only 71 frames in total. The third dataset, *Dataset_Aug*, is the augmented dataset by RFPose-GAN, which contains both the real samples and the synthetic samples for each of the five types of activities. The augmented dataset includes 284 frames (213 synthesized plus 71 sampled) for each type of activity. The test dataset includes real samples for all five types of activities mentioned above, but is distinct from any of the training datasets, so that the trained model can be realistically evaluated with new, unseen data.

B. Evaluation Results and Discussions

To showcase the benefits of the data augmentation approach, we train the same RFID-Pose model [4] with the augmented dataset and an original small dataset, respectively. The trained models will then be used for 3D human pose tracking using the same new RFID test data. Snapshots of testing results when the subject is walking are presented in Fig. 4 and Fig. 5. In both figures, the left-hand-side plot (blue) is the ground truth pose obtained by Kinect. while the right-hand-side plot (red) is the pose estimated using the trained RFID-Pose model. As shown in Fig. 4, the estimated pose using the model trained with augmented data is quite close to the ground truth. On the other hand, Fig. 5 shows that the estimated pose obtained using the model trained with the original small dataset, even though roughly mimicking the ground truth pose, exhibits some apparent differences from the ground truth, e.g., the estimated right arm and right leg. In fact, the estimated four limbs using the limited training dataset struggle to imitate the ground truth pose all the time. This example demonstrates that the model trained by the augmented dataset achieves a higher accuracy on estimating the joints positions, despite that the original dataset only consists of a limited amount of ground truth RFID data.

The performance of RFPose-GAN is also evaluated on estimating different types of poses. We use the Euclidean distances between the joint positions in the estimated 3D pose and that in the corresponding ground truth pose to indicate estimate error. The overall estimation error $\mathcal{E}_{all}$ used in our experimental evaluation is given by:

$$\mathcal{E}_{all} = \frac{1}{12}\sum_{n=1}^{12} ||\hat{P}_n - \dot{P}_n||, \tag{1}$$

where $\hat{P}_n$ represents the estimated position of joint n, $\dot{P}_n$ denotes the ground truth position collected by Kinect 2.0 for joint n, and $||\hat{P}_n - \dot{P}_n||$ is the Euclidean distance between the two 3D positions.

Fig. 6 shows the mean pose estimation error comparison for the three models that are respectively trained using the three datasets. A trained model with insufficient data usually performs poorly when tested on unseen data. The model trained using *Dataset_3Act* has the largest errors for all the five tested activities, which are all greater than 11.41cm. The model trained with *Dataset_5Act* performs relatively better, but its smallest mean error among the five activities is still 7.46cm (for walking). The poor performance is as expected due to the limited amount of data used for model training. With data augmentation, the accuracy of pose estimation is considerably improved across all the tested activities. The largest error of the model trained by *Dataset_Aug* is 7.84cm for squatting, which is lower than the minimum error of the other two models.

The efficacy of data augmentation can be further demonstrated by the Cumulative Distribution Function (CDF) curves

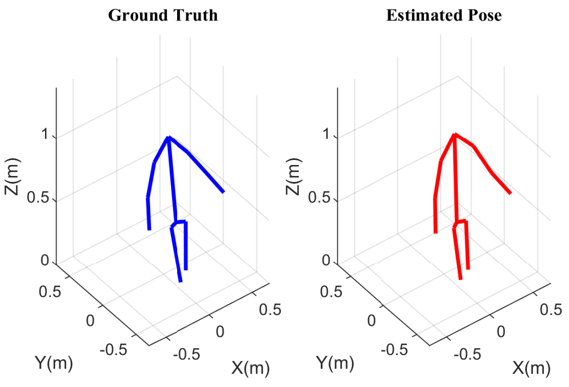

Fig. 4. Example of the estimated pose obtained by the trained model using the augmented dataset.

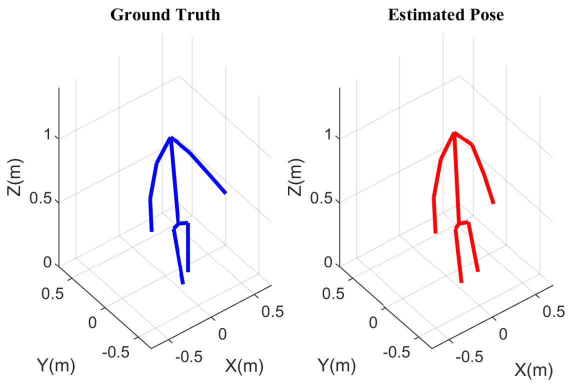

Fig. 5. Example of the estimated pose obtained by the trained model using the limited original dataset.

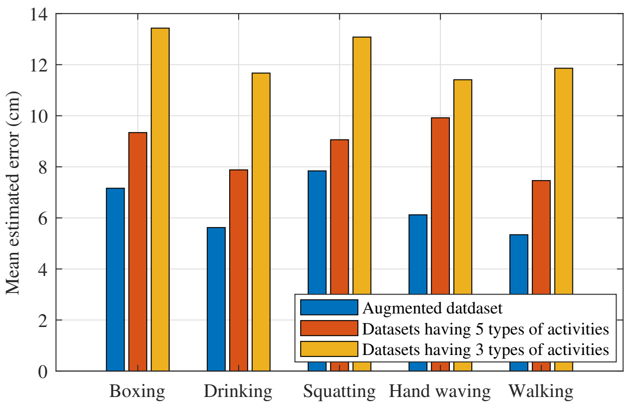

Fig. 6. Mean estimation errors errors of the three models trained with *Dataset_3Act*, *Dataset_5Act*, and *Dataset_Aug*, respectively.

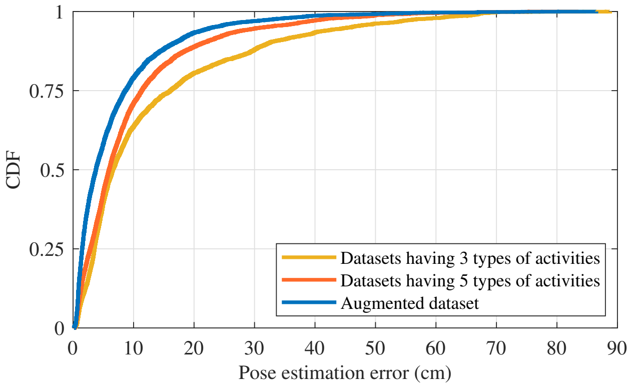

Fig. 7. CDF curves for estimation errors of the three models trained with *Dataset_3Act*, *Dataset_5Act*, and *Dataset_Aug*, respectively.

presented in Fig. 7. It can be seen that the median error is 3.77cm for the augmented model, 5.97cm for the model trained with *Dataset_5Act*, and 6.61cm for the model trained with *Dataset_3Act*. The largest error for the augmented model is significantly smaller than that of the other two models. These results validate that the proposed RFPose-GAN can effectively generate synthesized RFID data that are useful. The RFID pose model trained with the augmented dataset estimates human pose more accurately than the models that are trained with the unaugmented datasets.

IV. Conclusions

In this paper, we proposed RFPose-GAN for data augmentation for RFID-based 3D human pose tracking. Our approach was to augment the existing RFID dataset by generating synthesized samples paired with given pose samples using a GAN model. Through an experimental study, we demonstrated that the cost of data collection for vision-assisted RFID-based 3D human pose tracking can be greatly reduced with the proposed data augmentation approach, and the synthesized RFID data is useful for training the deep pose tracking model.

Acknowledgment

This work is supported in part by the NSF under Grants ECCS-1923163 and CNS-2107190.

References

[1] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” in *Proc. IEEE CVPR’17*, Honolulu, HI, July 2017, pp. 7291–7299.

[2] A. Sengupta, F. Jin, R. Zhang, and S. Cao, “mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs,” *IEEE Sensors J.*, vol. 20, no. 17, pp. 10 032–10 044, Sept. 2020.

[3] W. Jiang, H. Xue, C. Miao, S. Wang, S. Lin, C. Tian, S. Murali, H. Hu, Z. Sun, and L. Su, “Towards 3D human pose construction using WiFi,” in *Proc. ACM MobiCom’20*, London, UK, Sept. 2020, pp. 1–14.

[4] C. Yang, X. Wang, and S. Mao, “RFID-Pose: Vision-aided 3D human pose estimation with RFID,” *IEEE Transactions on Reliability*, vol. 70, no. 3, pp. 1218–1231, Sept. 2021.

[5] C. Yang, L. Wang, X. Wang, and S. Mao, “Environment adaptive RFID based 3D human pose tracking with a meta-learning approach,” *IEEE J. Radio Freq. Identif.*, to appear. DOI: 10.1109/JRFID.2022.3140256.

[6] C. Yang, X. Wang, and S. Mao, “Subject-adaptive skeleton tracking with RFID,” in *Proc. IEEE MSN’20*, Tokyo, Japan, Dec. 2020, pp. 599–606.

[7] M. Patel, X. Wang, and S. Mao, “Data augmentation with Conditional GAN for automatic modulation classification,” in *Proc. ACM Workshop on Wireless Security and Machine Learning (WiseML’20)*, Linz, Austria, July 2020, pp. 31–36.

[8] H. Shangguan and R. Mukundan, “3D human pose dataset augmentation using generative adversarial network,” in *Proc. 3rd Int. Conf. Graphics Signal Process.*, Hong Kong, China, June 2019, pp. 53–57.

[9] Y. Chen, Y. Tian, and M. He, “Monocular human pose estimation: A survey of deep learning-based methods,” *Elsevier Comput. Vision Image Understanding*, vol. 192, no. 3, p. 102897, Mar. 2020.

[10] R. Villegas, J. Yang, D. Ceylan, and H. Lee, “Neural kinematic networks for unsupervised motion retargetting,” in *Proc. IEEE CVPR’18*, Salt Lake City, UT, June 2018, pp. 8639–8648.