Demo Abstract: AIGC for RFID-based Human Activity Recognition

Ziqi Wang and Shiwen Mao
Dept. of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201 USA
Email: zzw0104@auburn.edu, smao@ieee.org

***Abstract*—The lack of sufficient radio frequency (RF) data constrains the performance of AI-empowered wireless communications, networking, and sensing research. RF data collection is more difficult and costly than other data types (e.g., text or image). To this end, we propose to exploit the strength of diffusion models on latent domains to generate super-realistic data for RF sensing applications. In this demo, we present a novel lightweight AIGC framework centered on latent domains, termed Activity Class Conditional Latent Diffusion Model (RFID-ACCLDM), for easy generation of large amounts of RF data at low cost, conditioned on activity class labels. We demonstrate the high performance of RFID-ACCLDM with RFID-based 3D human pose estimation and human activity recognition (HAR) model as representative downstream tasks.**

***Index Terms*—AIGC, Conditional Latent Diffusion Models, Human Activity Recognition (HAR), Internet of Things (IoT).**

I. Introduction

Deep learning has enabled great advances recently in the field of wireless communications and networking [1], but adequate model performance often requires a large amount of labeled RF training data. Furthermore, typical RF data possesses strong temporal, spectral, and spatial dependencies, rendering collecting RF datasets an extremely expensive task. Sometimes, a collected RF dataset may have limited value in a different setting due to the lack of diversity.

On the other hand, Artificial Intelligence-Generated Content (AIGC) has started a revolution in the machine learning field, where abundant high-quality text, image, or video data, along with their labels, can be generated with the typing of a prompt. A natural question is raised: *Can we leverage the power of AIGC to tackle wireless communication problems, especially to generate super-realistic RF data, while maintaining diversity*? Such AIGC RF data will greatly benefit many downstream wireless tasks, and in particular, RF sensing, such as creating a more robust human activity recognition (HAR) system that can be easily deployed in various diverse environments [2].

In this demo, we present an Activity Class Conditional Latent Diffusion Model (termed RFID-ACCLDM), a conditional latent diffusion model (CLDM) capable of generating super-realistic RFID sensing data of rich diversity, based on user input of desired human activity class labels [3]. To reduce the computational cost, while increasing the generative quality of diffusion models, we first train a recurrent variational autoencoder (R-VAE) to enable sampling latent representations that encapsulate the temporal dependency of the RF sensing data. Next, we leverage a CLDM to train the diffusion process on the RF latent dimensions.

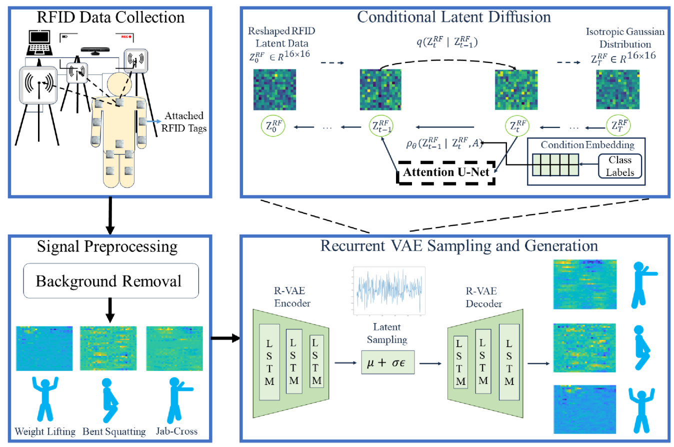

Fig. 1. Architecture of the RFID-ACCLDM framework.

II. System Overview

A. System Architecture

An overview of the system architecture is presented in Fig. 1, which comprises three major modules, including (i) RFID signal collection and processing, (ii) The R-VAE and CLDM architecture, and (iii) downstream tasks.

B. RFID Signal Preprocessing

Test subjects are first attached with RFID tags to their joints. RFID phase difference signals and 3D pose data are sampled simultaneously by an RFID reader and an Xbox Kinect, respectively. The sensing devices have different sampling frequencies, 110Hz for RFID and 30Hz for Kinect. Hence the sensory samples are downsampled to 7.5Hz for synchronization. Hampel filters are then used to remove background noise and DC components.

C. Latent RFID Activity Data Diffusion

We integrate Long Short-Term Memory (LSTM) units into the Variational Autoencoder to capture the temporal coherency in 2D time-variant RFID data with rich features and sample latent vectors with such time dependencies. The latent vectors are then fed into a diffusion model [4], which consists of two modules typically named "forward" and "reverse" diffusion process. The former utilizes a Markov chain with a fixed-variance scheduler to progressively contaminate the latent vectors with random Gaussian noise until eventually altering the data distribution into an Isotropical Gaussian distribution,

IEEE INFOCOM 2024 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) | 979-8-3503-8447-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/INFOCOMWKSHPS61880.2024.10620814

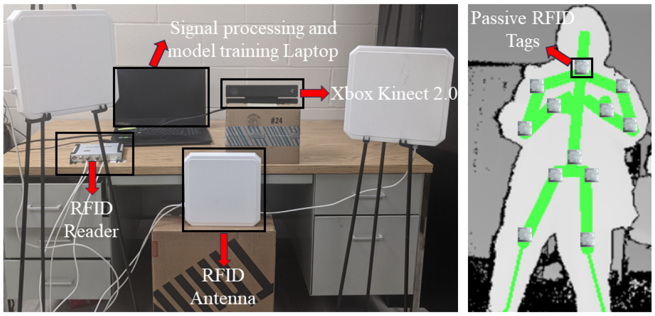

Fig. 2. An implementation of the proposed system and experimental setting.

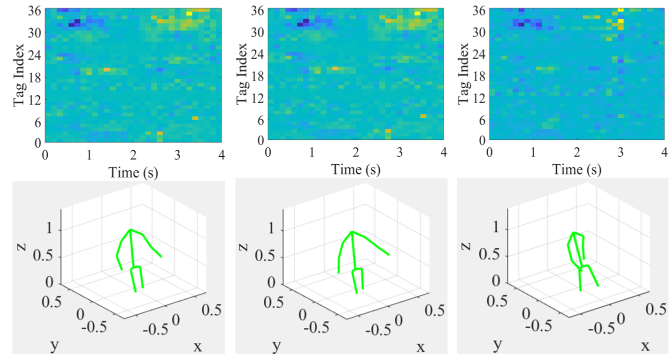

Fig. 3. A visual illustration of our AIGC generated RFID data: The top row is a visual comparison of generation quality between the real data (left) and the generated data (middle and right) in forms of images with scaled color for the waving up and down action. The bottom row are the animations of 3D walking pose estimated from our data by the RFID-Pose network [6] with a 0.8-second difference between the three animation video frames presented.

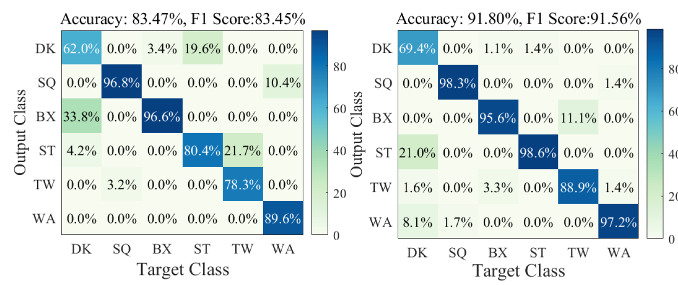

Fig. 4. confusion matrices of HAR obtained from CNN models trained on 32 minutes of real data (left) and 64 minutes of AIGC generated data (right).

while the latter leverages a Markov chain as well to learn the transitional kernels that can be parametrically modeled by a U-Net [5], to remove the Gaussian contamination. The activity class label embedding is incorporated into the U-Net throughout the diffusion training for generation of labeled samples. Finally, high-quality RFID data can be reconstructed from the latent representations, which highly resemble the "real" RFID data collected in a real deployment of the system. The detailed implementation can be seen in [3] .

III. Implementation and Evaluation

1) Hardware and Software Platforms: Three S9028PCR polarized antennas, one Impinj R420 reader, and ALN-9634 (HIGG-3) passive RFID tags make up an RFID-based implementation of the proposed framework, shown in Fig. 2. A Lenovo laptop with a GTX 1660 Ti GPU is used for the signal processing and model training tasks. We only use around 8 minutes of data for training both R-VAE and CLDM models. These data were captured from three test volunteers with similar body shapes. It takes 4 hours to train the R-VAE. As for the diffusion training, we use a linearly scaled variance β_t from $\beta_0 = 10^{-4}$ to $\beta_T = 0.02$. We also set the number of noising steps T to 1,000. A cyclical learning rate mechanism with the maximum learning rate set to 0.005 is used. The training takes 12 hours, Whereas the training on raw RFID data takes 16 hours. Our approach only takes 4 seconds to generate one sample, while the latter takes nearly 40 seconds.

2) Evaluation and Results: Test subjects perform distinctive activities in front of the sensing platforms, including drinking (DK), squatting (SQ), boxing (BX), standing still (ST), twisting (TW), and walking (WA). We first evaluate the generative ability of our system through visualization. The bottom row of Fig. 3 showcases that our generated RFID data presents fine-grained movement information that can be seamlessly mapped to 3D human pose animation. Besides the apparent anthropomorphically plausible posture, the 3D human poses exhibit natural temporal smoothness similar to real poses captured by camera-based devices. The top row of Fig. 3 demonstrates the impressive diversity of our generated RFID data, with the generated data on the right different from its real counterpart, while remaining fidelity. Next, we conduct a six-class HAR utilizing a simple and traditional Convolutional Neural Networks (CNNs) model. Test data include two different subjects at locations slightly different from where the training data was collected. Fig. 4 shows that, with the addition of 32 minutes of synthesized data, both metrics surpass the case of training with real data by a considerable gap of around 9.7% improvement. The superiority of our AIGC model is evident from the fact that it takes us only 24 minutes to generate this amount of synthesized data. Our working model can currently generate RF data of 9 distinct classes. An extended evaluation will be provided in the Journal.

Acknowledgment

This work is supported in part by the US NSF under Grants CNS-2107190, CNS-2319342, and IIS-2306789.

References

[1] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key technologies and open issues," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3072–3108, 4th Quarter 2019.

[2] Z. Wang and S. Mao, "AIGC for wireless data: The case of RFID-based human activity recognition," in *Proc. IEEE ICC 2024*, Denver, CO, June 2024, pp. 1–6.

[3] Z. Wang and S. Mao, "AIGC for RF sensing: The case of RFID-based human activity recognition," in *Proc. ICNC 2024*, Big Island, HI, Feb. 2024, pp. 1–6.

[4] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *arXiv preprint arxiv:2006.11239*, Dec. 2020.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI 2015*, Munich, Germany, Oct. 2015, pp. 234–241.

[6] C. Yang, Z. Wang, and S. Mao, "RFPose-GAN: Data augmentation for RFID based 3D human pose tracking," in *Proc. IEEE RFID-TA 2022*, Cagliari, Italy, Sept. 2022, pp. 138–141.