AIGC for Wireless Data: The Case of RFID-based Human Activity Recognition

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

***Abstract*—Although great advances have been made in machine learning (ML) based wireless communications and networking, the performance of most ML-based schemes is heavily dependent on the availability of large amounts of high quality radio frequency (RF) data, which are more challenging and costly to obtain than other forms of data. To address this challenge, we propose to leverage diffusion models to generate high quality RF data, and develop a novel lightweight AIGC model for RF sensing, termed RFID-ACCDM (Activity Class Conditional Diffusion Model). RFID-ACCDM can synthesize large amounts of RF data at low cost, conditioned on a particular activity class. The high quality of RFID-ACCDM generated data is demonstrated by metrics of Structural Similarity Index (SSIM) and Frechet Inception Distance (FID), as well as a representative downstream task of human activity recognition (HAR), where the model trained with sufficient synthesized data outperforms the model trained by real data.**

***Index Terms*—AIGC, Conditional diffusion, Data augmentation, human activity recognition, RFID sensing.**

I. Introduction

The recent decade has witnessed considerable advances in machine learning (ML) based wireless communications and networking [1]. However, the performance of most ML-based schemes is heavily dependent on the availability of large amounts of high quality radio frequency (RF) data. Compared to image or text, RF data has its unique features and high quality RF data is much harder to collect. First, the captured RF data is highly susceptible to the open-space channel; any change in transceiver location and the propagation environment may result in a new data domain. Second, RF data is also highly dependent on the frequency band, as well as the transceiver devices and protocols (e.g., waveforms). For instance, a 900 MHz RFID channel is fundamentally different from a 60 GHz millimeter wave channel. Third, the wireless channel is also time-varying: a WiFi channel during business hours would look much different from that in the midnight. Due to such spatial, spectral, and temporal dependencies, it is very costly to collect RF datasets, while a collected RF dataset may have limited use when the setting becomes different. Therefore, how to obtain high quality RF data with high diversity while at a low cost, would be the first barrier to overcome to make "ML/AI for wireless" successful.

Another trend in the past couple of years is artificial intelligence generated content (AIGC). Prominent products, such as ChatGPT, DALL-E, and Codex, are paving the way for Artificial General Intelligence (AGI). These applications generally use transformer and diffusion models as backbone, and are mostly used in the context of text-to-image generation or text-prompted AI agents. A natural question is "can we exploit AIGC to address wireless communication problems, and in particular, to generate RF data?" Generative Adversarial Networks (GANs) [2], as a relatively older generation of AIGC technology, have been explored for data augmentation over the years [3]–[5]. However, most works are only able to use synthesized data to boost the performance of the existing dataset via augmentation or fine-tuning [4]. The complications of wireless data, coupled with the difficulty of training a GAN model, usually result in synthesized data with low fidelity or low diversity. The simple and low-dimensional synthesized data would be of limited value for RF sensing applications such as human activity recognition (HAR) [6], [7].

To this end, there have been several recent works on 3D pose estimations in the computer vision (CV) domain [8]–[10], which utilized diffusion models to generate 3D *pose animation data* with great fidelity and diversity. Motivated by these interesting works, in this paper, we propose to go one-step further, to use diffusion models to generate high quality *RF data* for HAR. Specifically, we shall develop a novel lightweight AIGC model for RFID sensing, termed RFID-ACCDM (Activity Class Conditional Diffusion Model) to synthesize high quality RF data at low cost conditioned on a particular activity class. The proposed RFID-ACCDM system is illustrated in Fig. 1. As a representative example of downstream tasks, we also design an RFID sensing system, which queries the RFID tags attached to a test subject's joints to recognize human activity. The conditional diffusion based RFID-ACCDM system will generate large amounts of high fidelity, high diversity data at low cost for training the RFID sensing system, thus saving the huge efforts on collecting training RF data.

The main contributions made in this paper can be summarized as follows:

- To the best of our knowledge, this is the first work that harnesses the power of conditional diffusion models to generate RF data. The quality of the synthesized data, in terms of quantity, fidelity, and diversity, are all superior over existing approaches. More important, the proposed AIGC model only requires a small amount of real RF training data to be effective.
- We qualitatively demonstrate the performance of RFID-ACCDM through a visual comparison of its synthesized data with ground truth. Furthermore, we quantitatively show that our generated data is of high fidelity and diversity through metrics of Structural Similarity Index (SSIM) [11] and Frechet Inception Distance (FID) [12].

ICC 2024 - IEEE International Conference on Communications | 978-1-7281-9054-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICC51166.2024.10622401

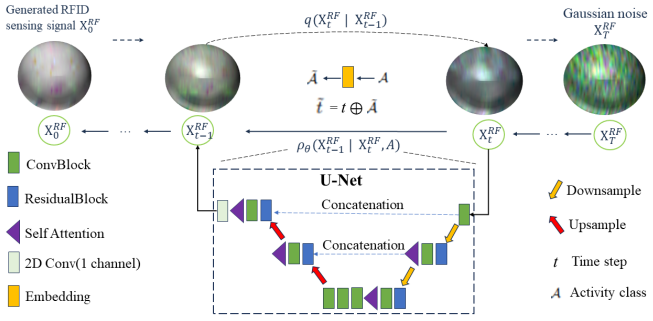

Figure 1. The procedure of conditional RF data generation with RFID-ACCDM. The reverse process p (see (4)) gradually converts random noises into plausible time series data, conditioned on embedded class labels. The structure of the noise predictor, the U-Net model, is illustrated in detail.

- Using a representative downstream task of HAR with RFID sensing, we demonstrate that our RFID-ACCDM generated data is highly effective in boosting the HAR performance without the need for real RF data.

In summary, we address two important problems with an *AIGC for Wireless approach*: how to avoid the high cost of collecting RF data, and how to synthesize RF data with high fidelity and high diversity for effective training of ML models.

The remainder of this paper is structured as follows. We first review related work in Section II. We then introduce the background of diffusion in Section III. Section IV describes the proposed system design and Section V presents our experimental study. Section VI summarizes this paper.

II. Related Works

AIGC applications based on diffusion have mostly been concentrated in the field of CV. The pioneering Diffusion Probabilistic Model (DPM) was applied to general medical image segmentation in [13], where superior performance of segmentation tasks have been demonstrated over state-of-the-art methods. In [14], the authors leveraged conditional diffusion models (CDM) for image-to-image translation, which outperformed GANs. Recently, it has been proven that diffusion models are also capable of generating continuous time-series data. The authors in [15] leveraged a Conditional Score-based Diffusion model to impute time-series healthcare and environmental data, which outperformed classic RNN-based models. A recent work [16] applied diffusion models to enable reliable 3D monocular pose estimation, to effectively reduce the inherent uncertainty and occlusion. RF data typically involves time frames and RF features in multiple dimensions. Since diffusion models are proficient with images and time-series data, they should also be a good fit for RF sensing tasks.

Various RF sensing applications have been developed for detecting human activities [6]. The impact of changes in the environment, user location and orientation, and user herself, tends to require large amounts of training data with high quality and diversity, in order to train models with good generalibility. To meet these requirements, GAN-based approaches have been explored recently [3], [5], [17]. For example, Li et al. [17] proposed the Amplitude-Feature Deep Convolutional GAN (AF-DCGAN) to generate additional channel state information (CSI) amplitude feature maps in order to reduce the effort of collecting WiFi fingerprints. However, Since CSI data is quite sensitive to environmental dynamics, any change in the indoor environment may result in a drop in location accuracy. Furthermore, most GANs can only synthesize additional data based on existing data with limited diversity. In [18], a multimodal GAN was proposed to deal with environmental changes. However, the multimodal system is rather complicated, consisting of two generators and one classification model. Overall, GANs have proved to be still effective for data augmentation in the wake of diffusion and transformer models, but generating useful and high-quality synthesized data tend to depend on complex procedures and mutlimodal systems.

III. Diffusion Preliminaries

This work follows the philosophy of the most prominent diffusion-based architecture proposed in [19]. The underlying idea is that the model can progressively improve its output through a series of small adjustments, ultimately yielding a high-quality sample. Diffusion models are based on a rather simple concept. They start with an input x_0 and slowly corrupt the input over a series of time steps (T) into a Gaussian distribution $\mathcal{N}(0, \mathbf{I})$ using fixed-variance-schedule defined Markov chain kernels, which is referred to as the *forward process*, given by [19]:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}). \tag{1}$$

At each time step t, Gaussian noise with a variance of β_t is added to x_{t-1}, resulting in a new latent variable x_t. $\mathbf{I}$ is the identity matrix, ensuring that each and every dimension of the multi-dimensional input have the same variance β_t. Therefore, the process, starting from input x_0 to x_T, can be tracked with $q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$. The sampling of x_t can then be admitted for any time step t in closed form, as [19]:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}), \tag{2}$$

where $\bar{\alpha}_t$ is given by $\prod_{\tau=0}^{t} \alpha_\tau$ and using β_t, α_t can be defined as $(1-\beta_t)$. We then obtain x_t using a reparameterization trick in a recursive manner, as:

$$\begin{aligned} x_t &= \sqrt{\alpha_t} \cdot x_{t-1} + \sqrt{1-\alpha_t} \cdot \epsilon_{t-1} \\ &= \sqrt{\alpha_t\alpha_{t-1}} \cdot x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}} \cdot \epsilon_{t-2} = \dots \\ &= \sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1-\bar{\alpha}_t} \cdot \epsilon_0, \end{aligned} \tag{3}$$

where $\epsilon_t \sim \mathcal{N}(0, \mathbf{I})$, $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{\tau=0}^{t} \alpha_\tau$.

On the other hand, the *reverse process* denoises the noisy inputs, after the forward diffusion process applies noise steps up to a certain point ($t \leq T$), to recover x_0. The reverse process is defined by the following Markov Chain [19]:

$$\begin{aligned} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) &= \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)) \\ p_\theta(\mathbf{x}_{0:T}) &= p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t). \end{aligned} \tag{4}$$

In [19], the authors showed that the reverse process can be trained by solving the optimization problem given below.

$$\min_\theta \ \mathbb{E}_{t,\epsilon,x_0} = \|(\epsilon - \epsilon_\theta(x_t, t))\|^2, \tag{5}$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ is the Gaussian noise added to the noisy input x_t, and ϵ_θ is a trainable denoising function that is often learned through a neural network such as U-Nets [20]. So one can interpret $\epsilon_\theta(x, t)$ as the noise vector estimated by the trained U-Net. This simplified training loss is made possible by the parameterization of $p_\theta(x_{t-1}|x_t)$, as:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} (\epsilon - \epsilon_\theta(x_t, t)) \right), \quad (6)$$

After training the diffusion model, high-quality samples x_0 can be obtained as given in (4).

IV. System Design

A. RFID Data Generation with Conditional Diffusion

In our prior work [6], we showed that by attaching RFID tags to human joints, RF information describing joint movements can be obtained, hence creating a complex high-dimensional data that can enable high-performance 3D human pose estimation and activity recognition. However, training the models requires large amounts of synchronized RFID and vision data, which is very costly to collect. Inspired by [16], where complicated 3D human animations represented by 3D joint coordinates are generated with high fidelity through diffusion models, we propose a conditional diffusion system termed RFID-ACCDM (i.e., RFID-based Activity Class Conditional Diffusion Model) to synthesize RFID data for various activity classes. Since the data are sampled when the test subject continuously repeats the activities, the data samples capture both short-range delicate movement information of human joints and long-range time-series information of movement trajectories. The proposed system can learn and leverage the inherent relationship between RFID data and 3D human movements to synthesize data with high fidelity for 3D human pose tracking and HAR.

Let x_t^{RF} denote the RFID data corresponding to a certain human activity at a random time step t, and $\mathcal{A}$ represent the class of human activity ranging from simple activities (e.g., standing still) to complex activities (e.g., boxing). We develop an RFID-sensing-specific reverse diffusion process and a supervised training method. The class condition $\mathcal{A}$ is taken as one of the inputs. The Markov chain for the reverse process of RFID-ACCDM is defined as follows.

$$p_\theta(\mathbf{x}_{t-1}^{RF}|\mathbf{x}_t^{RF}, \mathcal{A}) = \mathcal{N}(\mathbf{x}_{t-1}^{RF}; \boldsymbol{\mu}_\theta(\mathbf{x}_t^{RF}, t \,|\, \mathcal{A}), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t^{RF}, t \,|\, \mathcal{A}))$$
$$p_\theta(\mathbf{x}_{0:T}^{RF}|\mathcal{A}) = p(\mathbf{x}_T^{RF}) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}^{RF}|\mathbf{x}_t^{RF}, \mathcal{A}).$$

We then consider utilizing the following parameterization for ϵ_θ, which is different from (6) in that the class label of human activity $\mathcal{A}$ is taken as the condition in the newly defined conditional denoising function $\epsilon_\theta(\cdot)$, given by:

$$\mu_\theta(x_t^{RF}, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t^{RF} - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} (\epsilon - \epsilon_\theta(x_t^{RF}, t \,|\, \mathcal{A})) \right).$$

Finally, we formalize the training objective for the RFID-ACCDM system as a minimization problem, given by:

$$\min_\theta \ \mathcal{L}_\theta = \min_\theta \ \mathbb{E}_{t,\epsilon,x_0} = \left\| \left(\epsilon - \epsilon_\theta(x_t^{RF}, t \,|\, \mathcal{A}) \right) \right\|^2. \quad (7)$$

Algorithm 1 Training Procedure of RFID-ACCDM

Input: $\mathcal{A}$

repeat
  $\tilde{\mathcal{A}} = \text{Embedding}(\mathcal{A})$;
  $x_0^{RF} \sim q(x_0^{RF})$;
  $t \sim \text{Uniform}(1, 2, ..., T)$;
  $\epsilon \sim \mathcal{N}(0, \mathbf{I})$;
  $x_t^{RF} = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$;
  $\tilde{t} = t \bigoplus \tilde{\mathcal{A}}$;
  Take gradient descent step on
    $\nabla_\theta = \left\| \left(\epsilon - \epsilon_\theta(x_t^{RF}, \tilde{t}) \right) \right\|^2$;
until Convergence

Algorithm 2 Sampling Procedure of RFID-ACCDM

Input: $\mathcal{A}$

Sample $x_T \sim \mathcal{N}(0, \mathbf{I})$ with label $\mathcal{A}$;
for $t = T, ..., 1$ **do**
  if $t > 1$ **then**
    $z \sim \mathcal{N}(0, \mathbf{I})$;
  else
    $z = 0$;
  end if
  $x_{t-1}^{RF} = \frac{1}{\sqrt{\alpha_t}} \left(x_t^{RF} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t^{RF}, t \,|\, \tilde{\mathcal{A}}) \right)$;
end for
Return x_0^{RF};

Algorithm 1 and Algorithm 2 describe the training and sampling procedures of RFID-ACCDM, respectively.

B. U-Net for Denoising

As in [19], we adopt a U-Net [20] based on a wide ResNet for its desirable ability to facilitate the diffusion process. It takes in a noisy input at a particular time step and returns the predicted noise with the same size as the input data. The loss between the actually introduced noise ϵ and the predicted noise ϵ_θ is then used as the training objective. Loss minimization can be easily implemented via an MSE (mean squared error) function between ϵ_θ and ϵ at the current time step t in each training epoch. Sinusoidal positional encodings are applied to encode the time step t (also the noise level). The encodings help the network understand the particular time step it is at for each input within a batch during the diffusion process.

The activity class label $\mathcal{A}$ is embedded using a Pytorch package. This embedding layer works as an MLP (multilayer perceptron) layer, which is represented as a high-dimensional vector. MLP layers typically apply a linear transformation followed by a non-linear activation to obtain the embedding. The embedded label is next concatenated with time step t to integrate the class embedding into U-Net. This is possible because t is already implemented as a condition in the diffusion process. The resulting time step is denoted as $\tilde{t}$. The input RFID data, along with time step and class embedding, undergoes the standard encoder to decoder structure in the U-Net model. Our U-Net network has two downsampling operations realized by the `maxpooling2D` function in the

encoder, which reduce the spatial dimension to 16×3. At each downsampling step, we double the number of feature channels from 64 all the way to 256. Each downsampling operation is followed by a residual block and a convolutional block. The self attention mechanism is implemented right after the convolutional block with a multi-head self attention module to capture richer representations. The bottleneck (the middle block that keeps the feature size unchanged) has three convolutional blocks. A convolutional block has two 2D convolutional layers connected by a `GeLU` activation layer followed by a `GroupNorm` layer, with the final layer being another GroupNorm layer right after the second convolutional layer. A residual block is the same except for the addition of skip connections. The decoder is simply built with the reverse order as the encoder to recover the original input dimension, while concatenating the feature maps from the encoder. The basic convolution module has a kernel size of three. A deeper network can be used if there are larger numbers of joints. The complete conditional diffusion process along with a detailed structure of our implemented U-Net model are shown in Fig. 1.

V. Experimental Study

A. Implementation and Experiment Setting

We develop an RFID sensing system, as a representative downstream task, to evaluate the performance and benefits of our generative network. The system consists of an off-the-shelf Impinj R420 reader, passive ALN-9634 (HIGG-3) tags, and three S9028PCR polarized antennas. 12 RFID tags are attached to the test subject's joints (i.e., hip, neck, left upper leg, left knee, right upper leg, right knee, left shoulder, left arm, left forearm, right shoulder, right arm, and right forearm). An Lenovo Legion gaming laptop with an Nvidia GTX 1660 Ti GPU is used as the processor for signal processing and network training. The setup of the system is illustrated in Fig. 2. RFID data and vision pose data are collected simultaneously in front of the RFID system and an Xbox Kinect 2.0 device, when test subjects are performing different activities. The vision data will be used as labels for supervised training in the original, baseline system, where joint kinematics of vision data are transformed to the variations between RFID phase values from two consecutive time frames.

The frame rate of Kinect is 30 frames per second (fps), while the RFID data sampling rate is approximately 110 Hz. All data undergoes preprocessing and synchronization before being downsampled to 7.5 Hz. Throughout the experiment, both RFID phase variation data and 3D pose data are set to have a length of 8.53 seconds. A sliding window of 1.33 seconds is leveraged to create 4 basic data units with a length of 4 seconds. Their dimension is set to (30, 12, 3), with 30 being the frame number, 12 being the number of joints, and 3 being the number of antennas.

As for the diffusion training, we utilize a fixed linear β_t schedule from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$ with $T = 1,000$. Inspired by [21], a simple and elegant implementation of classifier free guidance is applied. In each epoch during training, we set the model to train unconditionally for 10% of the time; And in each epoch during sampling, we linearly interpolate from unconditional towards conditional sampling. This trick greatly enhances the stability of the model's ability to generate RFID phase variation data with corresponding classes, and simultaneously, the quality of the generated samples. A total of six RFID data files with a length of 64 frames per activity class are used as the U-Net's training data. These data are captured from three test volunteers with similar body shapes.

Figure 2. The configuration of the experimental system for RFID sensing.

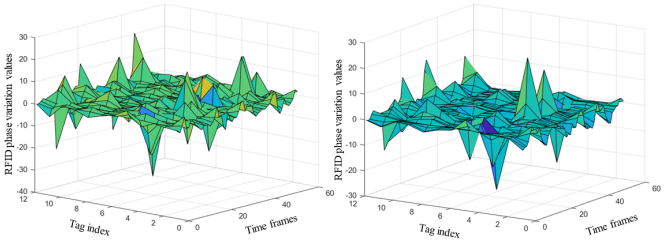

Figure 3. A visual comparison of generation quality between the real RFID data (left) and the generated RFID data (right) in the form of surface plots when the activity performed was walking.

B. Qualitative Results

In Fig. 3, we provide a visual comparison of the generated RFID phase variation data with the real RFID phase variation data captured for the activity. The surfaces are plotted in detail for a complicated activity of *walking* involving the movements of all limbs and torso. It can be seen that the generated RFID data, in the form of phase values, closely follows the movement trends along each joint and across time. There are small discrepancies between the generated and ground truth data, which help to enhance the diversity of the generated data.

C. Quantitative Results

In computer vision, the Structural Similarity Index (SSIM) [11] has been recognized as a useful metric to evaluate how similar the generated data is to the real data, since it not only measures the average intensity (luminance) and standard deviation (contrast), but also the details and general pattern of features inside an image using the structure index. For a regular image containing, say, a car or portrait, the structure index can locate the important features across all pixels, and similarly, the index is naturally suitable for evaluating the pattern of movement features across time frames and human joints. We choose a complicated activity of *boxing* to demonstrate the similarity in Fig. 4. The SSIM score is 0.65 in this case. From the SSIM map in Fig. 4(c), we can see that despite having matching patterns and periodicity, there is a certain amount of discrepancies between the real and generated data.

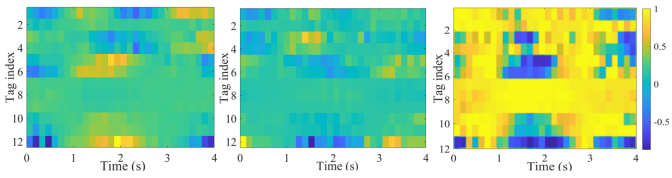

Figure 4. Generated (left) and real (middle) RFID data for boxing presented as images with scaled colors. On the right is the SSIM map showing the differences in each pixel between the generated and real data.

Table I
COMPARISON OF FID SCORES: RFID-ACCDM VS. RFPOSE-GAN

	Standing still	Waving	Walking
RFID-ACCDM (proposed)	8.7926	8.2465	20.6782
RFPose-GAN [5]	36.1981	32.2464	45.3412

Our proposed model also excels at generating high-quality RFID data with *great diversity*, instead of only generating homogeneous data similar to the training set (which yield high SSIM scores). Such diversity is critical for training a robust model, but cannot be accurately captured by SSIM. Also note that SSIM could introduce more bias as it focuses more on the evaluation of a single pair of real and generated data. Therefore, we also use the Frechet Inception Distance (FID) [12] to evaluate the distribution similarity between collections of generated and real RFID data. The FID score measures the distance between the feature vectors in the high dimensional latent space. The lower the FID score, the higher the fidelity of the generated RFID data as compared to the real data. Specifically, the FID score is defined as:

$$\varphi^2 = \|\mu_1 - \mu_2\|_2^2 + \mathrm{Tr}(Cov_1 + Cov_2 - 2\sqrt{Cov_1 \times Cov_2}),$$

where μ_1 and μ_2 refer to the feature-wise mean of the real and generated feature vectors, respectively, Cov_1 and Cov_2 are the covariance matrix of the real and generated feature vectors, respectively, and Tr is the trace linear algebra operation. A neural network, i.e., the `inceptionv3` model, is used to obtain the feature vectors between the two distributions.

Table I presents the superior FID scores achieved by the proposed model over our previous work RFPose-GAN [5]. RFPose-GAN deploys a supervised GAN that is capable of mapping one particular 3D pose data into its corresponding synthesized RFID data. GAN models are usually harder to train as they are in constant competition to synthesize data that rivals the distribution of real data. Synthesizing RF data for specific activities with minimal variations across time frames under interference and noise is highly challenging, hence the high FID scores of RFPose-GAN. The much lower FID scores achieved by the proposed RFID-ACCDM system are indicative of the high quality of the synthesized RFID data.

D. Human Activity Recognition Results

Perhaps the ultimate test for the synthesized RFID data is to examine how useful it is for training the ML model of a downstream task. We proceed to test the quality of our generated data utilizing a 6-activity-class RFID-based HAR system. A simple CNN model is employed for the classification task. There are 3 convolution layers, each followed by a dropout layer. A `maxpooling2D` layer is located after the second convolution layer. The convolution output is flattened and fed into a fully connected layer for the final accuracy calculation. Since a basic unit for activity classification has a length of 4 seconds, it takes about 30 minutes for the CNN model to achieve a modest 6-class HAR performance. This is why our proposed model becomes highly useful to mass-produce synthesized RFID data of high fidelity and diversity for any required activity. The test data are the collected ground truth data including two different subjects at locations slightly different from where the training data was collected.

Fig. 5 presents the confusion matrices for RFID-based human activity classification obtained by models trained by 32 minutes of real data (left), 32 minutes of RFID-ACCDM generated data, and 128 minutes of RFID-ACCDM generated data. When using the same amount of synthesized data, the accuracy and F1 score are both slightly lower than training with real data. However, with 128 minutes of synthesized data, both the accuracy and F1 score outperform the case of training with real data with considerable margins (about 8.3% improvements). This is an interesting finding, since the 128 minutes of synthesized data come at not much additional cost than running the RFID-ACCDM code a little bit longer.

Fig. 6 presents a comparison of F1 scores for progressively increased amounts of synthesized data using our proposed model at different training epochs. It can be seen that the F1 score is progressively improved as more synthesized data are used in model training. With 320 synthesized samples (i.e., 128 minutes), the F1 curve reaches 0.89 at 400 epochs. Compared with the case of using a modest amount of real data, the proposed approach achieves an approximately 8.3% gain in F1 score. Models trained on the generated data converge faster and achieve better results within the first 40 epochs. The model trained on real data suffers a more severe overfitting effect due to the lack of diversity, and converges at a slower pace. When there are 96 minutes or more of synthesized data, the F1 score exceeds that of 32 minutes of real data for all training epochs. This implies that more generated data help close the domain gap between real and generated data whereas RFPose-GAN synthesized data exhibit a rather large domain gap, causing performance issues. This will be further explored in the extended journal version of this work.

It is worth noting that the improved F1 scores are obtained by using pure synthesized data: *this is an AIGC for wireless sensing method, rather than a data augmentation method*. This experiment proves that the generated data by the proposed RFID-ACCDM method can effectively improve the accuracy of CNN-based HAR, further validating the fidelity and diversity of the AIGC RFID data using our model.

VI. CONCLUSIONS

In this paper, we addressed the RF data challenge with an AIGC for Wireless approach. The proposed RFID-ACCDM framework leverages a CDM to generate useful high-dimensional RFID sensing data conditioned on a class label. Through metrics of SSIM and FID, as well as a representative downstream task HAR, we demonstrated the high quality and usefulness of the synthesized data by the proposed RFID-

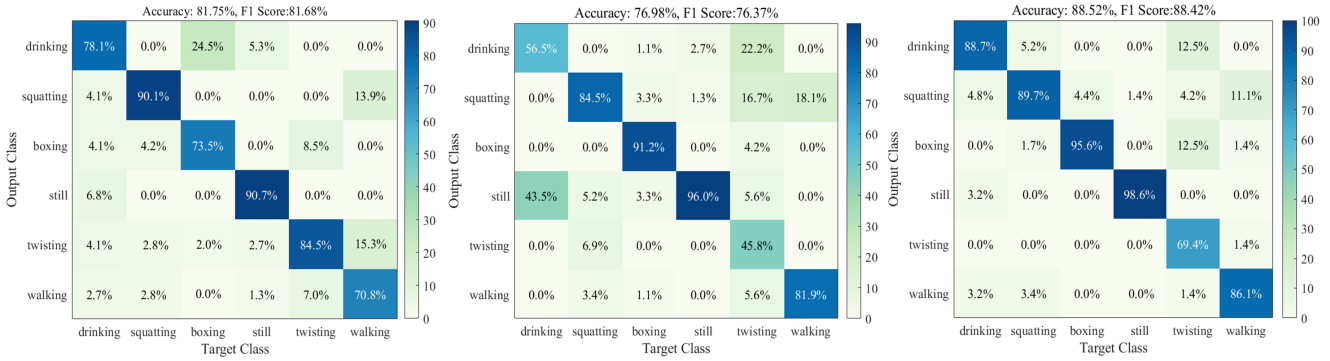

Figure 5. The confusion matrices obtained with the CNN model trained on 32 minutes of real data (left), 32 minutes of RFID-ACCDM generated data (middle), and 128 minutes of RFID-ACCDM generated data (right).

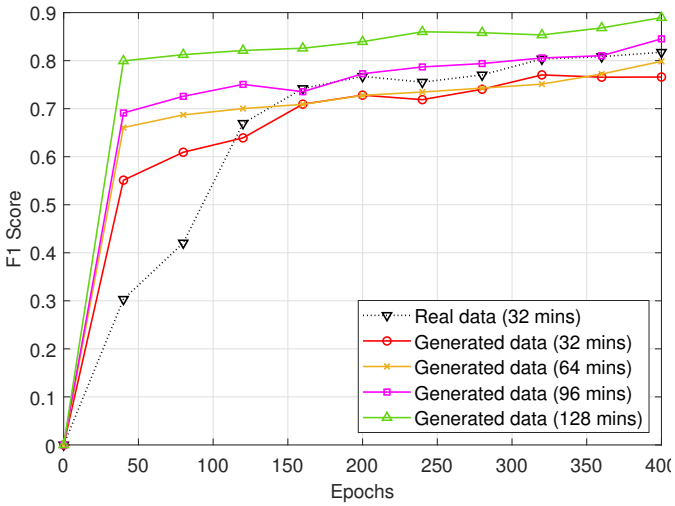

Figure 6. The F1 score of the classification with a progressively increase amount of generated data.

ACCDM system. The proposed *AIGC for wireless sensing approach* provided a compelling solution to the timely problems of how to avoid the high cost of RF data collection and how to synthesize high quality RF data.

Acknowledgements

This work is supported in part by the 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 Communications Surveys and Tutorials*, vol. 21, no. 4, pp. 3072–3108, Fourth Quarter 2019.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS 2014*, Montreal, Canada, Dec. 2014, pp. 2672–2680.

[3] M. Patel, X. Wang, and S. Mao, "Data augmentation with Conditional GAN for automatic modulation classification," in *Proc. ACM WiseML 2020*, Linz, Austria, July 2020, pp. 31–36.

[4] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-LSTM for human activity recognition using WiFi signal," *IEEE Internet of Things J.*, vol. 8, no. 6, pp. 4628–4641, Mar. 2021.

[5] Z. Wang, C. Yang, and S. Mao, "Data augmentation for RFID-based 3D human pose tracking," in *Proc. IEEE VTC-Fall 2022*, London, UK, Sept. 2022, pp. 1–2.

[6] 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.

[7] ——, "TARF: Technology-agnostic RF sensing for human activity recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 2, pp. 636–647, Feb. 2023.

[8] G. Tevet, S. Raab, B. Gordon, Y. Shafir, D. Cohen-Or, and A. H. Bermano, "Human motion diffusion model," in *Proc. ICLR 2023*, Kigali, Rwanda, May 2023, pp. 1–16.

[9] W. Shan, Z. Liu, X. Zhang, Z. Wang, K. Han, S. Wang, S. Ma, and W. Gao, "Diffusion-based 3D human pose estimation with multi-hypothesis aggregation," *arXiv preprint arXiv:2303.11579*, Aug. 2023. [Online]. Available: https://arxiv.org/abs/2303.11579

[10] C. Rommel, E. Valle, M. Chen, S. Khalfaoui, R. Marlet, M. Cord, and P. Perez, "DiffHPE: Robust, coherent 3D human pose lifting with diffusion," in *Proc. IEEE/CVF ICCV Workshops*, Paris, France, Oct. 2023, pp. 3220–3229.

[11] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transa. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[12] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. NIPS 2017*, Long Beach, CA, Dec. 2017, pp. 6629–6640.

[13] J. Wu, R. Fu, H. Fang, Y. Zhang, Y. Yang, H. Xiong, H. Liu, and Y. Xu, "MedSegDiff: Medical image segmentation with diffusion probabilistic model," in *Proc. Medical Imaging with Deep Learning 2023*, Nashville, TN, Mar. 2023.

[14] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi, "Palette: Image-to-image diffusion models," in *Proc. ACM SIGGRAPH 2022*, Vancouver, Canada, Aug. 2022, pp. 15:1–15:10.

[15] Y. Tashiro, J. Song, Y. Song, and S. Ermon, "CSDI: Conditional score-based diffusion models for probabilistic time series imputation," in *Proc. NeurIPS 2021*, Virtual Conference, Dec. 2021, pp. 1–13.

[16] J. Gong, L. G. Foo, Z. Fan, Q. Ke, H. Rahmani, and J. Liu, "DiffPose: Toward more reliable 3D pose estimation," *arXiv preprint arXiv:2211.16940*, Apr. 2023. [Online]. Available: https://arxiv.org/abs/2211.16940

[17] Q. Li, H. Qu, Z. Liu, N. Zhou, W. Sun, S. Sigg, and J. Li, "AF-DCGAN: Amplitude feature deep convolutional GAN for fingerprint construction in indoor localization systems," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 3, pp. 468–480, June 2021.

[18] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Multimodal CSI-based human activity recognition using GANs," *IEEE Internet of Things J.*, vol. 8, no. 24, pp. 17 345–17 355, Dec. 2021.

[19] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *arXiv preprint arxiv:2006.11239*, Dec. 2020. [Online]. Available: https://arxiv.org/abs/2006.11239

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Comput. Computer-Assisted Intervention 2015*, 2015, pp. 234–241.

[21] J. Ho and T. Salimans, "Classifier-free diffusion guidance," in *Proc. NeurIPS 2021 Workshops*, Virtual Conference, Dec. 2021, pp. 1–8.