A Study on Diffusion Modelling For Sensor-based Human Activity Recognition

1st Shuai Shao
Department of Computer Science
University of Warwick
Coventry, UK
Shuai.Shao.1@warwick.ac.uk

2nd Victor Sanchez
Department of Computer Science
University of Warwick
Coventry, UK
V.F.Sanchez-silva@warwick.ac.uk

***Abstract*—Human activity recognition (HAR) is a core research topic in mobile and wearable computing, and has been applied in many applications including biometrics, health monitoring and sports coaching. In recent years, researchers have focused more attention on sensor-based HAR due to the popularity of sensor devices. However, sensor-based HAR faces the challenge of limited data size caused by the high cost of data collection and labelling work, resulting in low performance for HAR tasks. Data transformation and generative adversarial network (GAN) have been proposed as data augmentation approaches to enrich sensor data, thereby addressing the problem of data size limitations. In this paper, we studied the effectiveness of diffusion-based generative models for generating synthetic sensor data as compared to the other data augmentation approaches in sensor-based HAR. In addition, UNet has been redesigned in order to improve the efficiency and practicality of diffusion modelling. Experiments on two public datasets showed the performance of diffusion modelling compared with different data augmentation methods, indicating the feasibility of synthetic sensor data generated using diffusion modelling.**

I. Introduction

Human activity recognition (HAR) involves using technology to identify and classify the physical movements of individuals, and has been applied in areas such as biometrics, healthcare, sports, and entertainment [1]–[4]. Sensor devices such as smartphones, smartwatches, and fitness trackers are becoming increasingly popular for tracking and monitoring a person's daily activity. A wide range of physiological and kinematic parameters can be measured by these devices, including heart rate, step count, movement speed, and a wide range of activity information. Sensor-based HAR has been gaining popularity in recent years for various purposes, for example, in biometrics to detect patterns of behaviour unique to individuals. As a form of biometric signature, these patterns can be used to identify or authenticate an individual, and can also be used to detect if a person is attempting to impersonate another by mimicking their movement patterns. Additionally, sensor-based HAR can be used in situations in which traditional biometric characteristics cannot be easily accessed, such as when a person is wearing gloves or masks, or when there is poor lighting or image quality. These cases can be addressed by sensor-based HAR, utilizing the movement patterns of the individual being identified.

979-8-3503-3607-8/23/$31.00 ©2023 IEEE

Using sensor devices for activity recognition is associated with challenges and limitations. These include issues related to sensor accuracy and data quality [5]. The capability of sensor-based HAR has been greatly improved by recent advances in machine learning (ML). In order to improve the model's performance, these methods require a large amount of data. However, it is often time-consuming and costly to collect and annotate sufficient sensor data. In addition, ML methods are sensitive to the quality of the data and may perform poorly if the data is noisy or contains biases. The difficulty of collecting sufficient quality sensor data has hindered research progress in the HAR field. Researchers have proposed some practical solutions to the problem of insufficient data in sensor-based HAR, such as data augmentation. In contrast to data transformation methods limited by the original data size, SensoryGans [6] used Generative Adversarial Network (GAN) as a data augmentation tool to generate additional synthetic sensor data. GAN was first introduced in [7] and has been proven successful in many areas, including computer vision [8], [9] and language processing [10], [11]. However, despite their success in other areas, GAN has not been widely used in the field of sensor-based HAR. GAN is difficult to train due to the large amount of data required and the constant competition between the generator and the discriminator, making training unstable and slow.

As compared to GAN, diffusion models have been shown to enhance data generation capabilities in computer vision, making them more appropriate for generating synthetic data and attracting more audience members. The Denoising Diffusion Probabilistic Model (DDPM) [12] has been popular in image generation within the computer vision community for some time. In this paper, we studied the diffusion modelling based on DDPM, and explored the possibility of applying the diffusion modelling to generate synthetic sensor data. Prior to this work, diffusion models had not been adapted for sensor-based HAR tasks. The main contributions of this work are as follows.

- To the best of our knowledge, this is the first study to compare diffusion modelling with other data augmentation approaches for sensor-based HAR.

- Redesigned the UNet model to make the diffusion modelling more applicable to sensor data.
- Visual analysis of the synthetic sensor data generated by diffusion modelling.

The remaining of this paper is organized as follows. In section II, we review the related work. In section III, we explain the DDPM in detail. In section IV, we introduce the method for applying diffusion modelling to sensor-based HAR, and the experiments are shown in section V. In section VI, the results of the experiments and some insights are discussed. Finally, section VII summarizes this work.

II. Related Work

A. Sensor-based Human Activity Recognition

Sensor-based HAR is a rapidly growing field with numerous applications in different areas. A wide range of approaches and techniques have been proposed and used for this purpose, the machine learning using manual feature engineering is the most common method for HAR tasks [4], [13]. However, feature engineering requires sufficient quantities of sample data [14].

In light of the difficulties in collecting and annotating sensor data, data augmentation is gaining more attention among researchers. Several methods have been proposed and used for augmenting sensor data. One method is to apply transformations to the data [15], such as scaling, jittering, rotation, and cropping, to create new data points. However, the original data size may limit the size of the new data that can be generated by the transformation method. Another method is to use synthetic data [6], [16], which is generated by a machine learning model. Synthetic data has the advantage of being able to simulate a wide range of scenarios and can be generated in large quantities. While synthetic data can produce significant amounts of data, existing GAN-based methods suffer from a number of shortcomings. In [6], since GANs are challenging to train, each activity may need its own GAN to generate related synthetic data. However, creating a separate GAN for each activity is not efficient and cannot be generalised in practice.

B. Diffusion Models

The diffusion model was first proposed in 2015 and inspired by Nonequilibrium Thermodynamics [17]. Currently, diffusion models have gained attention in the field of computer vision due to their ability to generate high-quality images while also preserving fine details. These models are based on the idea of using diffusion processes to denoise images and improve their quality. One significant research is DDPM [12] which has demonstrated its effectiveness for image generation. Following this, a number of variations of diffusion models have been proposed, including those based on image generation [18], image super-resolution [19], anomaly detection [20] and other similar topics. Although diffusion models have proven their effectiveness in the field of computer vision, it is still a blank for sensor-based HAR.

III. Denoising Diffusion Probabilistic Model

Diffusion refers to the process of spreading through time. In the context of DDPM, this means that the model attempts to smooth out the noise in the data by considering the values of nearby points in the time series. DDPMs have the advantage of being probabilistic models, which means that they can provide estimates of the uncertainty in their estimates of the underlying structure. This can be particularly useful for making predictions or for identifying patterns in the data that may not be immediately apparent.

In the forward process, the Gaussian noise is used to gradually degrade the samples in T steps, where the samples are from a real data distribution $\mathbf{x}_0$ ... $q(\mathbf{x})$, according to a variance schedule β_1 . . . β_T. And the noise samples are produced as $\mathbf{x}_1$ ... $\mathbf{x}_T$. Followed as (1):

$$\begin{aligned} q(\mathbf{x}_t|\mathbf{x}_{t-1}) &= \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \\ q(\mathbf{x}_{1:T}|\mathbf{x}_0) &= \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}) \end{aligned} \tag{1}$$

By using the reparameterization trick, we can sample $\mathbf{x_t}$ at any arbitrary time step t by Equation (2). Let $\alpha_t = 1-\beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t}\alpha_i$, we can have:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}) \tag{2}$$

The reverse diffusion process $p_\theta(\mathbf{x}_{0:T})$ gradually denoises from a Gaussian noise input $\mathbf{x_T} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and allows generating new data samples. It follows the reverse steps in Equation (3):

$$\begin{aligned} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) &= \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \mathbf{\Sigma}_\theta(\mathbf{x}_t, t)), \\ p_\theta(\mathbf{x}_{0:T}) &= p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \end{aligned} \tag{3}$$

In training, the usual variational bound on negative log-likelihood is optimized as follows:

$$\begin{aligned} \mathbb{E}\left[-\log p_\theta\left(\mathbf{x}_0\right)\right] \leq \mathbb{E}_q\left[-\log \frac{p_\theta\left(\mathbf{x}_{0:T}\right)}{q\left(\mathbf{x}_{1:T} \mid \mathbf{x}_0\right)}\right] = \\ \mathbb{E}_q\left[-\log p\left(\mathbf{x}_T\right) - \sum \log \frac{p_\theta\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t\right)}{q\left(\mathbf{x}_t \mid \mathbf{x}_{t-1}\right)}\right] =: L \end{aligned} \tag{4}$$

And L in Equation (4) can be written as:

$$L = \mathbb{E}_q\left[\begin{array}{c} \underbrace{D_{\mathbf{KL}}\left(q\left(\mathbf{x}_T\mathbf{x}_0\right)\|p\left(\mathbf{x}_T\right)\right)}_{L_T} + \\ \sum_{t>1}\underbrace{D_{\mathbf{KL}}\left(q\left(\mathbf{x}_{t-1}\mathbf{x}_t, \mathbf{x}_0\right)\|p_\theta\left(\mathbf{x}_{t-1}\mathbf{x}_t\right)\right)}_{L_{t-1}} \\ \underbrace{-\log p_\theta\left(\mathbf{x}_0\mathbf{x}_1\right)}_{L_0} \end{array}\right] \tag{5}$$

where KL refers to the Kullback-Leibler divergence.

The reverse conditional probability is tractable when conditioned on $\mathbf{x}_0$:

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\mu}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t\mathbf{I}) \tag{6}$$

In [12], the covariance is set as a constant and calculated $\mu_\theta(\mathbf{x}_t, t)$ as a function of noise:

$$\mu_\theta = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, t)) \tag{7}$$

where $\alpha_t = 1 - \beta_t, \bar{\alpha_t} = \prod_{i \le t} \alpha_t$, and $\epsilon_\theta(\mathbf{x}_t, t)$ is defined as a function approximator which can predict ϵ from $\mathbf{x}_t$. The simplified mean-squared error between $\epsilon_\theta(\mathbf{x}_t, t)$ and ϵ along with time t is defined as:

$$L(\theta) = \mathbb{E}_{t,\mathbf{x}_0,\epsilon}\left[\left\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)\right\|^2\right] \tag{8}$$

IV. Diffusion Modelling For Sensor-based HAR

In this paper, we studied diffusion modelling based on DDPM, which is a generative model for reconstructing latent signals from noisy observations for sensor-based HAR. We followed the same process of DDPM, which consists of two stages, forward diffusion and backward denoising. Instead of feeding image data, we modified the input layer to accept sensor signal data. We applied the sliding window technique [21] to segment sensor signals into frames to capture proper activity information. Each frame represents 1s sensor signals. Then, diffusion and denoising processes can be performed using frames of sensor data.

The UNet [22] was introduced in DDPM to represent the reverse process, which improved the quality of image synthesis. UNet is initially designed for image segmentation. It has a U-shaped architecture that consists of an encoder network to downsample the input image, and a decoder network to upsample the encoded image back to the original size. The traditional UNet is not suitable for processing sensor data directly because of the significant difference between sensor data and images. In addition, UNet has a deep network for downsampling and upsampling, which is ineffective for sensor data with fewer channels or short frame lengths.

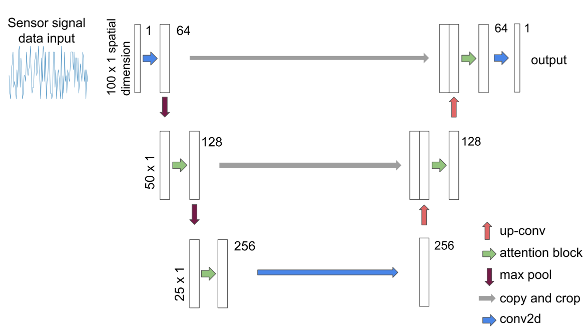

Fig. 1. The proposed UNet architecture (example for 100x1 sensor data)

Instead of using a traditional UNet, we redesigned the UNet shown as Fig. 1, which can take sensor data and be more efficient when applied in HAR tasks. We redesigned the input layer of UNet to accept sensor data frames with fewer channels (e.g. single channel) and different frame lengths. Due to the simpler nature of sensor data, we restructured the downsampling and upsampling layers. In the downsampling layer, we proposed using a self-attention layer followed by a 2x1 max pooling. In the upsampling layer, we proposed using a self-attention layer followed by 2D convolutions. The network was optimized for HAR tasks by reducing the contracting path (left side) and expansive path (right side) to avoid over-featuring the sensor data with fewer channels or short frame lengths. In the training process, we embedded class labels so that a unified framework could be developed to handle a variety of activities. Lastly, we generated synthetic sensor data in the form of frames using diffusion modelling.

V. Experiments

A. Dataset

We choose two public datasets for our experiments. The HASC2010corpus (HASC) [23] consists of seven subjects and six activities. Sample data includes accelerometer values for three axes. In PAMAP2 [24], twelve activities were recorded across nine subjects using IMUs (including accelerometers, gyroscopes, and magnetometers) attached to hands, chests, and ankles. Our data is recombined based on different activities in the datasets (accelerometer data only in HASC, and wrist-worn accelerometer data only in PAMAP2), and three typical activities (Stay, Walk, and Jog) are chosen to represent the original data. The hold-out method is used to split the dataset into training and testing data. In order to prevent information leakage, generative models are trained using the training data only to generate the synthetic data, and using testing data to verify the quality of generated data by evaluating the performance of selected ML models. Since machine learning models may be preferred when testing the quality of generated data due to their simplicity, interpretability, efficiency, and their statistical understanding. In order to eliminate direction effects from accelerometers, we preprocessed the values using Euclidean norm as $X = \sqrt{{a_x}^2 + {a_y}^2 + {a_z}^2}$, where a_x, a_y, a_z are accelerometer values in three different directions respectively. By applying sliding window techniques, we were able to obtain the data of the appropriate size. Each dataset had a frame length of 1s, and all frames were normalised by $Z = \frac{X-\mu}{\sigma}$, where μ and σ are the training data's mean and standard deviation respectively. Afterwards, normalised data is used for all analyses.

B. Result

In order to verify the reliability of the diffusion modelling for sensor-based HAR, four different data augmentation methods are chosen for comparison. These methods include data transformation and data generation.

We evaluated synthetic data using Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), k-Nearest Neighbor (KNN) and Support Vector Machine (SVM). The parameters of these classifiers were all selected by grid search. We selected fourteen statistical features extracted by sliding windows from raw and synthetic data. The features include mean, standard deviation, min, max, 50th and 75th percentile,

TABLE I
THE MEAN F1 RESULT ON REAL VS SYNTHETIC DATA

		LR	DT	RF	KNN	SVM
HASC						
Real Data	Baseline	0.898	0.901	0.903	0.901	0.899
Data Transformation	Data Scaling [15]	0.901	0.903	0.904	0.899	0.901
	Data Jittering [15]	0.899	0.899	0.903	0.901	0.900
Synthetic Data	SensoryGANs [6]	0.908	0.910	0.912	0.908	0.911
	ActivityGAN [16]	0.896	0.902	0.901	0.899	0.903
	Diffusion Modelling	**0.916**	**0.918**	**0.923**	**0.920**	**0.919**
PAMAP2						
Real Data	Baseline	0.935	0.938	0.942	0.936	0.937
Data Transformation	Data Scaling [15]	0.936	0.940	0.944	0.937	0.936
	Data Jittering [15]	0.935	0.938	0.943	0.936	0.935
Synthetic Data	SensoryGANs [6]	0.945	0.947	0.948	0.946	0.945
	ActivityGAN [16]	0.945	0.946	0.946	0.947	0.944
	Diffusion Modelling	**0.948**	**0.949**	**0.953**	**0.950**	**0.949**

TABLE II
THE MEAN F1 RESULT FOR TRAINING ON REAL DATA AND TESTING ON SYNTHETIC DATA IN HASC

		LR	DT	RF	KNN	SVM
Training/Testing (real data/real data)	Baseline	0.898	0.901	0.903	0.901	0.899
Training/Testing (real data/synthetic data)	SensoryGANs [6]	0.900	0.901	0.902	0.899	0.900
	ActivityGAN [16]	0.897	0.898	0.901	0.900	0.899
	Diffusion Modelling	0.901	0.902	0.903	0.903	0.902

kurtosis, skewness, standard error of the mean, median, interquartile range, range of values, median absolute deviation, and count below mean.

We measured performance using the mean F1-score for all experiments, which is defined as $\bar{F}_1 = \frac{1}{K}\sum_{k=1}^{K}\frac{2\mathbf{TP}_k}{2\mathbf{TP}_k+\mathbf{FP}_k+\mathbf{FN}_k}$, where K stands for the number of classes (activities). $\mathbf{TP}_k$, $\mathbf{FP}_k$, $\mathbf{FN}_k$ denote the number of true positive, false positive, and false negative predictions, respectively.

We compared diffusion modelling with other data augmentation approaches in Table I. In light of the simplicity of the activity and the distinction between activities, the raw data allows the model to learn certain features for classification, thereby weakening the role of data transformation in performance improvement. Therefore, data transformation methods perform at the same level as baseline results. On the other hand, synthetic data can improve the model's performance. Since synthetic data can be labelled with no errors or ambiguities, and it can be generated with an even distribution of classes, these properties can help avoid bias when training models. We can see that SensoryGANs has a better performance than ActivityGAN in both datasets, particularly in HASC where the mean F1 score is increased by 1% for all ML methods. In spite of SensoryGANs achieving better results, it is not practical because it requires a separate GAN for each activity. While ActivityGAN can generate all activities from a unified architecture, it fails to improve performance. Conversely, we found that the data generated by diffusion modelling can obtain satisfactory performance gains compared to data transformation and GANs. In HASC, the mean F1 score of all ML models can be improved by more than 1%, and reached 2% for RF and SVM. As opposed to SensoryGANs, diffusion modelling has a unified architecture, making it more efficient and practical. In PAMAP2, where the activities are more standardized, the diffusion modelling can still improve the mean F1 score by more than 1% compared with the baseline.

1) More testing on synthetic data: The quality of synthetic data was further explored to confound the model's perception of the data. As seen in Table II, there are two groups of testing: The first group is the same baseline experiment as in Table I, and the second group is an additional verification for synthetic data, where the model trained using real data, and then tested using synthetic data. This allows us to verify whether the model is capable of distinguishing between real and synthetic data.

From the result in Table II, SensoryGANs has similar results as baseline, while ActivityGAN has a slight decrease, suggesting these GANs can generate accurate synthetic sensor data. A better classification performance is obtained with synthetic data generated by diffusion modelling compared with baseline and other GANs, which indicates that diffusion modelling has the ability to generate higher-quality data.

These experiments have shown that the data generated by diffusion modelling may have similar features to real data, which can prove the quality of synthetic data. However, testing on synthetic data may not be useful in a real scenario, the result in Table I suggests that the synthetic data could contribute to the training of the model.

C. Synthetic Data Visualization

To assess the quality of the synthetic data, we selected three activities (Stay, Walk, and Jog) which can be distinguished by visual observation. In this paper, we compared the amplitudes of real and synthetic data to visualize the degree of similarity. Inspired by the approaches in [6], we applied the synthetic sensor data in two different visualization methods: local visualization and global visualization.

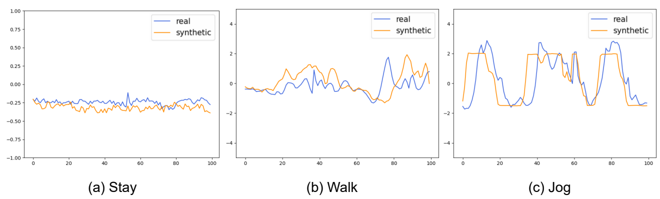

Fig. 2. Local visualization of real and synthetic data in HASC dataset

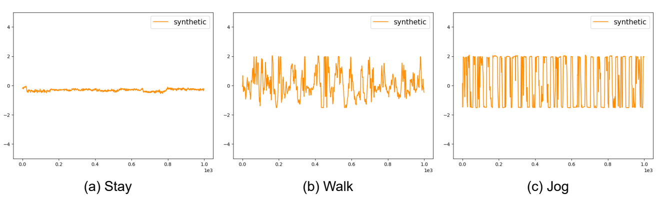

Fig. 3. Global visualization of synthetic data in HASC dataset

1) Local Visualization: In order to investigate whether local amplitude trends are similar between the synthetic and real data, we first accessed training data for one sliding window, as shown in Fig. 2, the blue line is real data and the orange line is synthetic data. As can be seen from the amplitude trends observed for all three activities, the synthetic data is highly similar to the real data. As a result of these observations, it is clear that the synthetic data generated by the diffusion modelling demonstrate a satisfactory representation of real data.

Additionally, synthetic data offers other benefits beyond its ability to mimic real data to a high degree. While the sliding window technique is widely accepted in sensor-based HAR tasks, it suffers from the problem that the window may contain activities that are not matched with the labels. Nevertheless, synthetic data generated by diffusion modelling could correct the problem and generate accurate data that matches the label.

2) Global Visual Evaluation: In addition, we evaluated the synthetic data from a global perspective illustrated in Fig. 3. Based on the results, we can clearly identify synthetic data for different activities. While the stay activity in Fig. 3 (a) shows the smoothest trend, the walk and jog in Fig. 3 (b, c) have much more varying amplitude, but we can also clearly identify them based on their frequency. Compared to real data, the

TABLE III
COMPARISON OF UNET AND REDESIGNED UNET

	Parameter Size	Time for generating one-second frame data
UNet	14988865	3s
Resigned UNet	3794369	1s

generated data tend to be more regular and does not fluctuate as much as real data, which allows the data to better reflect the activity label.

D. Further Experiments

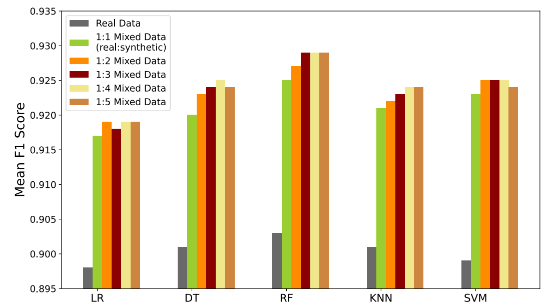

Fig. 4. The performance on HASC for different ratios of mixed data

1) The effect of training samples: Mixing real data with synthetic data not only increased the size of the data, but also provided comprehensive training samples. Synthetic data can be used to augment real data, by generating additional examples that are similar to the real data but differ in certain ways, which can help to improve the generalization of the model. To explore the relationship between mixed data ratio and performance, we tested mixing real data with synthetic data with ratios of 1:1, 1:2, 1:3, 1:4 and 1:5. From Fig. 4, we can see that the model's performance improves as data increases. The 1:1 mixed data has a significant performance gain compared with the real data baseline, and the best performance occurs with 1:4 mixed data for most cases. After a certain amount of mixed data from the 1:4 ratio, the performance gradually begins to level off.

2) The effect of UNet architecture: The UNet was redesigned to work better for sensor-based HAR tasks. As well as changing the input layer to accept sensor data, we also reduced the contracting path and expansive path, yielding a more efficient model with 75% fewer parameters and 67% less generation time (for one-second frame data) while keeping the same quality of synthetic data. Due to the nature of the simplicity of sensor data compared with the image, the simplified structure of UNet is more suitable for sensor-based HAR tasks.

VI. DISCUSSION

This paper demonstrates the potential of diffusion modelling in generating synthetic sensor data for human activity recognition (HAR) tasks. By specifying desired properties, such as noise levels and variability, diffusion modelling can generate predictable and consistent synthetic data, making it easier to train models. However, this regularity can also limit the ability of synthetic data to capture the complexity and variability of real-world data. In general, real data is preferred for training models, as it better reflects the diversity of real-world conditions. Synthetic data, however, can be a useful supplement shown in Fig. 4 when real data is scarce or expensive.

In this study, we utilized diffusion modelling to verify its feasibility for sensor-based HAR tasks, choosing three simple activities and single-channel sensor data. Our findings suggest that diffusion modelling outperformed other data augmentation methods, owing to the high quality of the synthetic data. Nonetheless, the approach faces challenges when dealing with complex activities or multi-channel sensor data. Therefore, further research is essential to improve the efficacy of diffusion modelling for sensor-based HAR tasks. Our study serves as a preliminary exploration of the potential of diffusion modelling for HAR, highlighting the need for more comprehensive research to deepen our understanding of this approach in the context of sensor-based HAR.

VII. Conclusion

In this paper, we studied diffusion modelling compared with other data augmentation approaches for sensor-based HAR. Based on three identifiable activities, diffusion modelling can generate high-quality sensor data, effectively providing a new direction for data augmentation in the HAR field. We demonstrated the usability and recognizability of the synthetic data through different visualization methods, illustrating the improved performance of the generated data in two different datasets. It is expected that sensor-based HAR research will benefit from the development of more diffusion models, especially in situations with limited resources.

References

[1] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living," *IEEE Access*, vol. 7, pp. 133 190–133 202, 2019.

[2] N. Hu, S. Su, C. Tang, and L. Wang, "Wearable-sensors Based Activity Recognition for Smart Human Healthcare Using Internet of Things," *2020 International Wireless Communications and Mobile Computing, IWCMC 2020*, pp. 1909–1915, 6 2020.

[3] KhanAftab, NicholsonJames, and PlötzThomas, "Activity Recognition for Quality Assessment of Batting Shots in Cricket using a Hierarchical Representation," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–31, 9 2017. [Online]. Available: https://dl.acm.org/doi/10.1145/3130927

[4] K. Sirlantzis, M. Mentzelopoulos, and A. Protopsaltis, "Human Computer Interaction using gestures for mobile devices and serious games: A review sotirios spanogianopoulos," *Proceedings of 2014 International Conference on Interactive Mobile Communication Technologies and Learning, IMCL 2014*, pp. 310–314, 1 2015.

[5] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep Learning for Sensor-based Human Activity Recognition," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, 5 2021. [Online]. Available: https://dl.acm.org/doi/10.1145/3447744

[6] J. Wang, Y. Chen, Y. Gu, Y. Xiao, and H. Pan, "SensoryGANs: An Effective Generative Adversarial Framework for Sensor-based Human Activity Recognition," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2018-July, 10 2018.

[7] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[8] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3722–3731.

[9] M. Zhang, K. Teck Ma, J. Hwee Lim, Q. Zhao, and J. Feng, "Deep future gaze: Gaze anticipation on egocentric videos using adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4372–4381.

[10] D. Croce, G. Castellucci, and R. Basili, "GAN-BERT: Generative Adversarial Learning for Robust Text Classification with a Bunch of Labeled Examples," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, 7 2020, pp. 2114–2119.

[11] Y. Yu, A. Srivastava, and S. Canales, "Conditional LSTM-GAN for Melody Generation from Lyrics," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 1, 4 2021.

[12] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851.

[13] S. Sani, N. Wiratunga, S. Massie, and K. Cooper, "kNN Sampling for Personalised Human Activity Recognition," in *Case-Based Reasoning Research and Development*, D. W. Aha and J. Lieber, Eds. Cham: Springer International Publishing, 2017, pp. 330–344.

[14] T. Plötz, N. Y. Hammerla, and P. Olivier, "Feature Learning for Activity Recognition in Ubiquitous Computing," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, ser. IJCAI'11. AAAI Press, 2011, pp. 1729–1734.

[15] T. T. Um, F. M. J. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, and D. Kulić, "Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring Using Convolutional Neural Networks," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ser. ICMI '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 216–220.

[16] X. Li, J. Luo, and R. Younes, "ActivityGAN: Generative Adversarial Networks for Data Augmentation in Sensor-Based Human Activity Recognition," in *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, ser. UbiComp-ISWC '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 249–254.

[17] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep Unsupervised Learning using Nonequilibrium Thermodynamics," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, 1 2015, pp. 2256–2265.

[18] P. Dhariwal and A. Nichol, "Diffusion Models Beat GANs on Image Synthesis," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 8780–8794.

[19] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 684–10 695.

[20] J. Wyatt, A. Leach, S. M. Schmon, and C. G. Willcocks, "AnoDDPM: Anomaly Detection With Denoising Diffusion Probabilistic Models Using Simplex Noise," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 650–656.

[21] A. Bulling, U. Blanke, and B. Schiele, "A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors," *ACM Comput. Surv.*, vol. 46, no. 3, 1 2014.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[23] N. Ogawa, K. Kaji, and N. Kawaguchi, “Effects of number of subjects on activity recognition-findings from hasc2010corpus,” in *International Workshop on Frontiers in Activity Recognition using Pervasive Sensing (IWFAR2011)*, 2011, pp. 48–51.

[24] A. Reiss and D. Stricker, “Introducing a new benchmarked dataset for activity monitoring,” *Proceedings - International Symposium on Wearable Computers, ISWC*, pp. 108–109, 2012.