SensoryGANs: An Effective Generative Adversarial Framework for Sensor-based Human Activity Recognition

Jiwei Wang*†‡, Yiqiang Chen*†, Yang Gu*†, Yunlong Xiao*†, Haonan Pan*†,
*Beijing Key Laboratory of Mobile Computing and Pervasive Device, Beijing, China
†Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
‡University of Chinese Academy of Sciences, Beijing, China
Email:{wangjiwei, yqchen, guyang, xiaoyunlong, panhaonan}@ict.ac.cn

***Abstract*—This study focuses on improving the performance of human activity recognition when a small number of sensor data are available under some special practical scenarios and resource-limited environments, such as some high-risk projects, anomaly monitoring and actual tactical scenarios. The Human Activity Recognition (HAR) based on wearable sensors is an attractive research topic in machine learning and ubiquitous computing over the last few decades, and has extremely practicality in health surveillance, medical assistance, personalized services, etc. However, with the limitation of sensor sampling rate, sustainability, deployment, and other restricted conditions, it is difficult to collect enough and resultful sensor data anywhere, anytime. Therefore, the HAR based on wearable sensors always faces the challenges of the low-data regime under some practical scenarios, which leads to a low accuracy of activity recognition and needs to be solved urgently. Currently, the Generative Adversarial Networks (GANs) provide a powerful method for training resultful generative models that could generate very convincing verisimilar images. The framework of GANs and its variants shed many lights on improving the performance of HAR.**

In this paper, we propose a new generative adversarial networks framework called SensoryGANs that can effectively generate available sensor data used for HAR. To the best of our knowledge, SensoryGANs is the first unbroken generative adversarial networks applied in generating sensor data in the HAR research field. Firstly, we tried exploring and devising three activity-special GANs models for three human daily activities. Secondly, these specific models are trained with the guidance of unbroken vanilla GANs. Thirdly, the trained generators from adversarial optimization process are used to generate synthetic sensor data. Finally, the synthetic sensor data from SensoryGANs are used to enrich the original authentic sensor datasets, which can improve the performance of target activity recognition model. Meanwhile, we propose three visual evaluation methods for assessing synthetic sensor data produced by the trained generators in SensoryGANs models. Experimental results show that SensoryGANs models have the capability of capturing the implicit distribution of real sensor data of human activity, and then the synthetic sensor data generated by SensoryGANs models have a potential for improving human activity recognition.

I. Introduction

The Human Activity Recognition(HAR) with wearable sensors has always been an active research area in pervasive and ubiquitous computing more than one decades[1] and has been grabbing considerable research attentions from a pretty wide range of researchers[2][3][4][5][6]. The HAR has immense potential in large numbers of application areas, such as health-care, entertainment, medical rehabilitation, elderly monitoring, physical therapy, fitness trackers, sleep quality monitoring, military, security monitoring and skill assessment under some special practical scenarios. The HAR can benefit researchers from better understand and analyze subjects' behavior. In the research field of HAR, sensor data collection is the foundation and headstream of the whole work. The key to enabling a high HAR accuracy is collecting enough effective sensor data of human activity. However, most of HAR researchers mainly focus on the optimization of classification methods rather than considering more high-quality sensor data. In some special circumstances, deploying enough sensors and obtaining sufficient sensor data are often time consuming and expensive. Even sometimes, data availability and data quality are poor due to the nonstandard action of subjects. These limitations will lead to unfavorable situations of inadequate sample information and incomplete data, that is, small-sample problem.

In view of the above problems, Our main objective is to research three questions:

(1) Can the GANs framework be employed to generate verisimilar sensor data of human activity as effective as generating images.

(2) Can those synthetic sensor data from GANs models enable high accuracy rate of human activity recognition.

(3) How can we evaluate the performance of GANs models for generating sensor data of human activity.

This paper explores an effective generative adversarial framework for generating synthetic sensor data, named SensoryGANs. It is inspired by the thought of GANs and can be used to guide the generator to produce more realistic sensor data of human activity. The GANs framework was firstly proposed by Ian Goodfellow et al. in 2014[7]. A generator model and a discriminator model both built by multilayer perceptrons are the basic modules of vanilla GANs. The goal of GANs is estimating generative models that can capture the distribution of real data with the adversarial assistance of a paired discriminator based on min-max game theory. After the birth

978-1-5090-6014-6/18/$31.00

of GANs, a great many variants of GANs have been widely researched to generate effective synthetic samples, such as image generation[7], image inpainting[8], image translation[9], super-resolution[10], image de-occlusion[11], natural language generation[12], text generation[13], etc. Though the powerful learning capabilities have gained great success in many fields, GANs have not yet been adapted to generating sensor data. As far as we known, three difficulties are hampering the development of GANs for generating sensor data. Firstly, the framework of GANs was designed for generating images which have different distribution space compared with sensor data. A uni-model distribution may be enough for GANs to learn the image, but not enough for GANs to represent a complex distribution of sensor data which may be multivalued at any given time. Secondly, sensor data are more sensitive to the back-propagating gradients. Thirdly, by far instability of training process is hard to overcome completely.

Generally, traditional machine learning based on manual feature extraction is the most commonly method for recognizing human activity. Deep learning models with automatic feature extraction are also widely adapted to this community[14]. However, feature learning relies on the availability of sufficient quantities of sample data[15]. Therefore, small-sample problem severely limits the potential contribution of human activity recognition technologies. In fact, most prior works on human activity recognition focused on the classification performances and paid less attention to the importance of preparing and generating ideal sensor data. To further improve the performance of HAR, efficient data generation method should be researched. Therefore, we have explored GANs framework which is a hot topic in deep learning community.

In the paper, we focus on generating effective sensor data of human activity with the availability of small amounts of original sensor data to improve the performance of human activity recognition. Therefore, we attempt to introduce the framework of GANs to solve this problem. To the best of our knowledge, SensoryGANs firstly adopt the complete framework of GANs to human activity recognition community based on sensor data. Though SensoryGANs framework is a complete framework of GANs family, it is very elementary work for generating effective sensor data of human activity. We have explored different kinds of neural networks to construct the SensoryGANs models and selected preferable models for certain activities respectively. In other words, the current version of SensoryGANs is not a single network model but multiple particular network models aiming at different activities. This diverse setting reveals that not only the RNNs but also CNNs are appropriate for learning the distribution of sensor data.

The main contribution of this paper is as follows:

(1) We have explored a kind of new-type application for the GANs framework. To the best of our knowledge, SensoryGANs is the first research of adapting unbroken GANs framework to generate sensor data for human activity recognition. The training processes do not depend on large amounts of training data, and the stability of training process can be guaranteed by SensoryGANs models.

(2) Three visual evaluation metrics are proposed to assess the quality of synthetic sensor data from SensoryGANs models.

(3) Through generating effective sensor data of human activity, SensoryGAN enable better HAR machine learning models without collecting plentiful of authentic data from realistic scenes or vast manual label work. Synthetic sensor data from SensoryGANs can effective improving the accuracy of human activity recognition.

The remaining of this paper is organized as follows. In section II, we will review previous related works on the generative adversarial networks, human activity recognition and a most similar work of generating sensor data. In section III, we will discuss the basic knowledge of GANs. In section IV, we will present the detail of the proposed SensorGANs models. In section V, experiments are deployed to evaluate the performance of SensoryGANs models. In section VI, we will give a conclusion of this paper and introduce the future works of SensoryGANs based on the shortages of current works.

II. Related Work

A. Generative Adversarial Networks

Firstly, the original GANs[7] was proposed to generate plausible fake images approximating real images in low-resolution, such as MNIST, TFD, CIFAR-10. The global optimality and convergence of GANs have been analyzed theoretically. And the potential of GANs has also been qualitatively and quantitatively evaluated. Many straightforward extensions of GANs have been demonstrated and led one of the most potential research directions. Then, inspired by this work, many variants of GANs have been explored and proposed, such as CGANs [16], LAPGANs[17], ACGANs[18], DCGANs[19], Progressive GANs[20]. The goal of these work is generating more and more verisimilar samples with higher resolution, simultaneously speeding and stabilizing the training process. At the same time, GANs framework is also explored and expanded to many other applications, such as resolving the finer texture details[10], image inpainting [8], language generation[12], text generation[21], natural language generation[13].

B. Human Activity Recognition

In the field of pervasive and ubiquitous computing, the Human Activity Recognition(HAR) based on wearable sensors has been discussed for more than one decade since it was firstly researched by F. Foerster et al.(1999) [1]. And then, large numbers of traditional machine learning methods and deep learning methods were adapted to classify human activities and have gained favorable recognition accuracy[2][22][3][4][5][6], but more preferable results rely heavily on well-designed and hand-crafted feature extraction techniques. Consequently, some works tried to explore and design more effective features to further improve the performance of HAR[23][24]. At the same time, it is well known that deep learning is a very powerful method which is effective for automatic feature extraction. Therefore, with the development of deep learning, deep classification models are also introduced to replace hand-crafted feature extraction for recognizing human activities and perform very well[14]. Some researchers also adopted multimodal techniques to improve the performance of HAR[23][25][26] with a fusion of different channels of sensor data. The HAR was useful in some practical areas, such as medical rehabilitation, health monitoring, security monitoring, entertainment and some military scenarios. Therefore, researchers have always been pursuing better technology to achieve more convincing recognition accuracy. Nevertheless, most works just focus on the optimization of classification methods and extracting more effective features that could represent sensor data[15]. However, few research works focus on the sensor data itself. In this paper, from another point of view, we aim to improve the quality of sensor data itself and try unbroken GANs models to achieve augmentation of original sensor data of human activity.

C. Sensor Data Generation

The sensor data of human activity are sequential time series that can be represented by $\{x_1, x_2, ..., x_n\}$. For sensor data generation, RNN networks are more competent because a hidden internal state memory in each RNN unit can be updated according to the previous state and new input value. However, it suffers from vanishing gradient problem and exploding gradient problem. Training GANs with recurrent neural networks (RNNs) are more challenging, mostly due to that sensor data have more than one value at any given step. More and more generative models adapt Long Short Term Memory(LSTM) networks and Gated Recurrent Units(GRU) to overcome gradient problems. Moustafa et al.[27] firstly tried to use the idea of GANs framework to train the LSTM-based generator to produce sensor data, but their SenseGen was half-baked GANs framework. Both the generator and the discriminator in their SenseGen were trained separately, that is, the training process of the generator in SenseGen was not based on the back-propagating gradient from the discriminator. SenseGen was, in essence, a straight generative model for synthesizing sensor data. In this paper, we focus on generating sensor data of human activity through unabridged GANs models for improving the accuracy of HAR.

III. Generative Adversarial Networks

A classical GANs framework is composed of a generator $G(z)$ and a discriminator $D(x)$, where z is random noise. The generator $G(z)$ tries to generate more and more verisimilar data to 'fools' the discriminator $D(x)$, while the discriminator $D(x)$ aims to tell apart the fake data from the real data. These two adversarial opponents are optimized to overpower each other and play a zero-sum game (also called the min-max game) in the whole training process. The random noises $z \in \mathbb{R}^N$ (usually normal distribution or gaussian distribution) are provided as the input of the generator $G(z)$. And then, the generator $G(z)$ will generate synthetic data, $\tilde{x} = G(z)$. The real data x and fake data $\tilde{x}$ will be both fed to the discriminator $D(x)$, and then the discriminator $D(x)$ will output a scalar which represents the probability of input data are from the real data distribution $P(x)$ rather than the generator $G(z)$. The two adversarial players are optimized by the adversarial training process. The value function of this adversarial process is as follows(GANs learn the generator G(z) and the discriminator D(x) by solving Nash equilibrium problem):

$$\begin{aligned} &\underset{G}{min}\underset{D}{max}V(D,G) = \\ &\mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \end{aligned} \tag{1}$$

where $P_z(z)$ is the distribution of random noises (uniform distribution in most GANs at the early phase, $P_z(z) = \mathcal{U}(0,1)$. Both the generator $G(z)$ and the discriminator $D(x)$ in original GANs[7] are built by multilayer perceptrons. They are both trained using stochastic gradient descent(SGD) according to the Equation 1.

From the perspective of Generator $G(z)$:

$$\underset{G}{min}V_G(D,G) = \underset{G}{min}(\mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]) \tag{2}$$

(1) $x_G = G(z)$ represents that the generator is modelled to transforms a random vector z into target sample x_G.

(2) $p_{data}(x_G)$ is maximized for training G (The probability that the generated samples belong to the distribution of real data).

(3) $p_z(z)$ is a fixed, easy sample prior distribution that GANs assumed.

From the perspective of Discriminator $D(x)$:

$$\begin{aligned} &\underset{D}{max}V_D(D,G) = \\ &\mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \end{aligned} \tag{3}$$

(1) GANs framework uses a sigmoid neuron at the last layer of Discriminator $D(x)$, so its output is in [0,1].

(2) The discriminator tries to assign a high value(the upper limit is 1) to real data, while assigning a low value(the low limit is 0) to fake data from the generator.

In fact, although some GANs frameworks aim at learning a discriminator[28], the original GANs and most of its variants are a class of methods for learning generative models based on game theory. This paper also targets at an effective generative model that can output synthetic sensor data with high availability quality. We mainly focus the application of GANs models rather than the optimization and stability of GANs framework in this paper.

IV. SensoryGANs

We dub our generative adversarial networks framework as SensoryGANs. It is an unbroken GANs framework that contains a generator and a corresponding discriminator. Because this is a very preliminary work that tries generative adversarial nets framework to generate sensor data of human activity, we construct particular SensoryGANs models for three specific human activities respectively. As illustrated in Figure 1, we build SensoryGANs for activity stay fundamentally based on the 1-dimensional convolutional networks, and the fully connected networks. And we construct models for walk activity mainly based on the bi-directional LSTM networks,

TABLE I: The basic modules for specific SensoryGANs models

	Stay	Walk	Jog
Generator	1-dimension CNN layers fully connected layers	bi-directional LSTM layers	LSTM layers
Discriminator	1-dimension CNN layers fully connected layers	1-dimension CNN layers fully connected layers	bi-directional LSTM layers fully connected layers

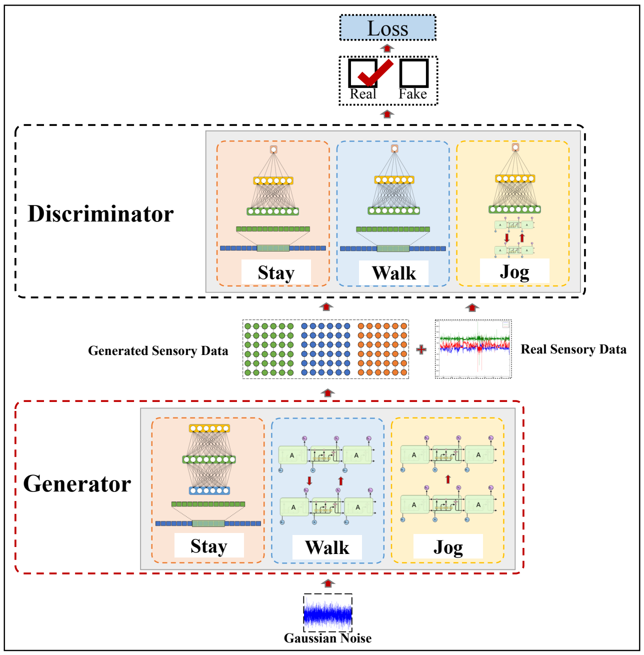


Fig. 1: The framework of SensoryGANs models

the 1-dimensional convolutional networks and fully connected networks. And we build models for jog activity principally based on the LSTM networks, the bi-directional LSTM networks and the fully connected networks. We try to construct a unified SensoryGANs model that could accommodate various distributions of different human activities. But the performance of unified SensoryGANs model for learning the different distributions of sensor data of human activity is not optimistic. In practice, we find that these personalized models designed for each activity can efficiently capture the implicit distribution of original real sensor data. In the future, we will explore new approaches to construct a convincing unified SensoryGANs model that is universally applicable to all kinds of sensor data for different human activities.

A. The Generative Models for Sensor Data

The target of training SensoryGANs is to learn a generator that can generate synthetic sensor data which will be used for recognizing human activity[22]. In another word, The goal of the generator is to learn a distribution $q(x)$ that can approximate the real distribution $p(x)$ accurately. Sensor data are time series from the real world distribution $p(x)$ that can be denoted as $X = \{x_1, x_2, ..., x_n\}$. Similarly, the synthetic sensor data from the generator $G(z)$ can be denoted as $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n\}$. Due to the dynamic temporal property of sensor data, we prefer to use RNN to process these time series. In practice, LSTM is adapted to construct the generator to capture the temporal characteristic and to avoid the serious vanishing gradient problem in RNN. The generator yields a sequence of sensor data by successively feeding preceding predictions to the following LSTM cells. At last layer of the generator, a series of synthetic sensor data are produced. In the training process, these fake sensor data from the generator are fed to the discriminator. While, in the generating process, the fake sensor data from the generator are used to improve the performance of the classifier for recognizing human activities.

In practice, we find that not only the LSTM but also 1-dimension CNN can learn the distribution of sensor data of human activity. For example, the learning capability of 1-dimension CNN for stay activities is superior to that of LSTM. So we construct different SensoryGANs models with a different kind of neural network as basic modules and select specific models with the superior performance for various activities. The optimum necessary modules for generators in SensoryGANs models for different activities are shown in Table I.

B. The Discriminative Models for Sensor Data

The discriminator is an assistant for improving the performance of the generator in the adversarial training process. It helps SensoryGANs to train the generator by back-propagating the gradient. Real sensor data will be given high values(the upper limit is 1), and synthetic sensor data from the generator will be given low values(the lower limit is 0). To make full use of the temporal features of sensor data, the discriminator of SensoryGANs also adapts LSTM and 1-dimension CNN as the basic function module. As typical phenomena in GANs, the discriminator is likely to overpower the generator, leading to the so-called vanishing gradient problem. Our strategy to solve this problem is to weaken the discriminator. In practice, for training the generator adequately, we trained the generator three times than the discriminator on each batch of sensor data. The optimum necessary modules for discriminators in SensoryGANs models for different activities are shown in Table I.

C. Pseudocode of Algorithm

Though all the SensoryGANs models contain three different models for three activities, both the whole training and generating process are similar. The pseudocode of SensoryGANs is as Algorithm 1. Because the generator in SensoryGANs models is our primary target, so we trained the generator

Algorithm 1 Pseudocode of SensoryGANs

Input: (1) random noises z; (2) real sensor data x
Output: synthetic sensor data $\tilde{x}$

The training and generating processes for three different activites are same
for the number of iterations **do**
  for the number of batches **do**
    Sample a batch of real sensor data $\{x_1, x_2, ..., x_n\}$
    Sample a batch of synthetic sensor data $\{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n\}$
    Update the discriminator with real and synthetic sensor data by Equation 3
    for k steps **do**
      Sample random noises from the Gaussian distribution, $z \sim N(\mu, \sigma^2)$
      Use random noises and gradients from the fixed discriminator to update the generator by Equation 2
for the number of sensor samples that need to be generated **do**
  Sample random noises from the Gaussian distribution, $z \sim N(\mu, \sigma^2)$
  Use the trained generator to output synthetic sensor data
return synthetic sensor data

three times than training the discriminator. This may lead to the instability of the discriminator. We adopted the Gaussian distribution to provide the generator with random noise. At each training epoch, the real sensor data and the synthetic data from the generator are concatenated to feed the discriminator. The synthetic data will be assigned higher value approximating 1, and the real data will be assigned lower value approximating 0. This reverse label assignment will be useful for the stability of the whole training processes.

V. Experiment

A. Dataset

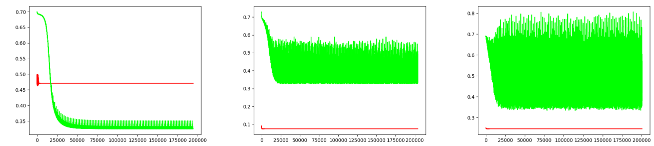
The public dataset adapted for training and evaluating SensoryGANs is an open human activity dataset, HASC2010corpus[22]. There are seven subjects and six activities for each subject. Every piece of data contains timestamp and 3-axis values of an accelerometer. We recombine these raw data based on six activities and prefer three typical behaviours as original data. To achieve better results and eliminate the effects of different directions, we feed consultant accelerometer values calculated by Equation 4 to SensoryGANs.

$$a_c = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{4}$$

where a_c is the consultant values of accelerometer, and a_x,a_y,a_z are accelerator values in three different directions respectively.

B. Training Processes of SensoryGANs

The principal goal of training GANs is to find a Nash equilibrium to a two-player adversarial game[29]. The stability



(a) Stay (b) Walk (c) Jog

Fig. 2: Training loss of generator and discriminator

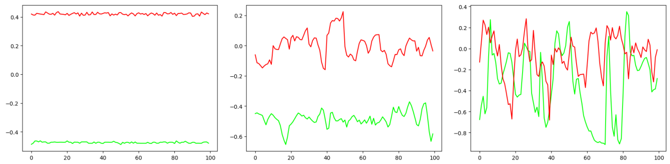
of training loss of the generator and discriminator shows that these two opponents can achieve dynamic equilibrium. Because this is the preliminary work using unbroken GANs framework for generating effective sensor data of human activity, we just use three typical activities with just accelerometer data. And in practice, we explore different kinds of neural network models to construct SensoryGANs and discover that 1-dimension convolutional networks are more suited for learning the stay activity, and LSTM[30] is suited for learning the activities of walk and jog. As illustrated in Figure 2, all the generators(red line) in different SensoryGANs models achieve stability at last, but the discriminators(green line) fluctuate severely. There are two possible reasons for this instability of discriminator. Firstly, we train the generators three times than training the discriminators. This may cause inadequate training for the discriminator. Secondly, the inherent changing features of sensor data sampled from different activities aggravate this phenomenon. As shown in Figure 2, the trend of fluctuations are consistent with rangeability of different activities(stay$<$walk$<$jog).

From the perspective of the generators, all of them in different SensoryGANs models can achieve stability at last. The training losses of the generators are ideal. This preliminary result accords with our final objective that generator can learn the true distribution of real sensor data and output synthetic sensor data approximating real sensor data steadily. In the future work, we will improve our SensoryGANs models to solve the instability of training the discriminators and achieve the Nash equilibrium between the generator and the discriminator.

C. Visual Evaluations

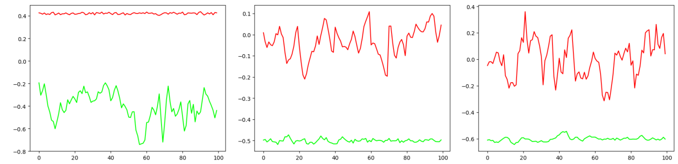
In the GANs community, the most common way for evaluating the quality of generated data is that letting human observers review them based on their prior knowledge. Considering the temporal features of sensor data of human activity, we proposed three visual evaluation methods for assessing the generated sensor data from SensoryGANs. Firstly, the Local Visual Evaluation is used to observe the performance of the generator in the training process. Secondly, the Global Visual Evaluation is used for observing the quality of synthetic data from the well-trained generator. Thirdly, the Memory Independence Visual Evaluation is used to evaluate the ability of avoiding overfitting problem.

1) Local Visual Evaluation: The Local Visual Evaluation is used to assess the training process of the generator in



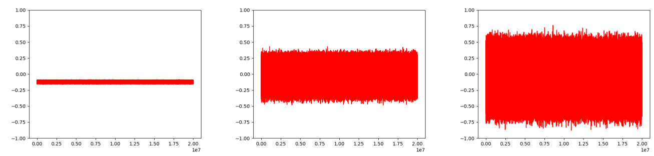
(a) Stay (b) Walk (c) Jog

Fig. 3: Local visual evaluation of generators in training process



(a) Stay (b) Walk (c) Jog

Fig. 5: Memory independence visual evaluation of trained generators



(a) Stay (b) Walk (c) Jog

Fig. 4: Global visual evaluation of the trained generators

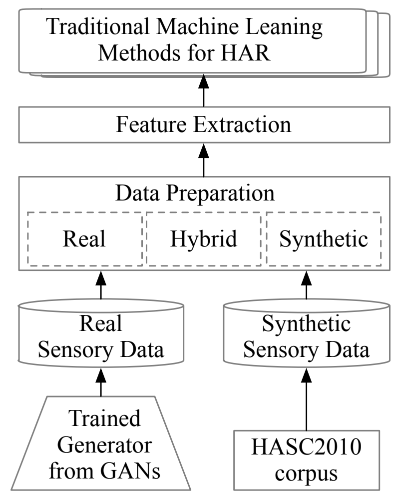


Fig. 6: Flow chart of evaluation process for the usability of SensoryGANs

SensoryGANs models. As illustrated in Figure 3, our SensoryGANs can learn the detail of raw sensor data for different activities(stay, walk, jog), where the red lines indicate the generated sensor data from the generator in the training process, while the green lines indicate the real sensor data. With the successive training of SensoryGANs models, the generators are improving the capability of capturing the distribution of the real sensor data of human activity. Though the performance of SensoryGANs model for jogging activity is not ideal, we will explore more effective SensoyGANs models for this activity in the future work.

2) Global Visual Evaluation: The Global Visual Evaluation is used to assess the quality of these synthetic sensor data from the well-trained generator. As illustrated in Figure 4, the value ranges of these synthetic sensor data for different activities are obvious. The synthetic stay sensor data are very gentle, while the synthetic walk and jog sensor data have some peaks, especially jog activity. These results are in line with the features of real-world activities. On the whole, these synthetic sensor data are too perfect, but we can find obvious differences between these three types of synthetic data from the visual view. In the future work, we will try to set more effective evaluation metrics and improve our SensoryGANs models.

3) Memory Independence Visual Evaluation: The memory independence metric is used to evaluate the capability that the SensoryGANs models can effectively avoid overfitting problem when learning the implicit distribution. As illustrated in Figure 5, SensoryGANs models can avoid overfitting problem when learning from limited data. In other words, SensoryGANs can generate sensor data without just simply remembering raw data. For example, time-series segmentation for human activity sensor data usually leads to impurity problem that one kind of activity sensor data may be mixed with more or less other contiguous activities sensor data, but our SensoryGANs can generate synthetic data without containing other two activities. The red line represents synthetic sensor data from SensoryGANs model. The green line denotes bad data mixed in each activity. All the SensoryGANs models can effectively avoid learning those impurities. This rational learning ability, to some extent, explains that our SensoryGANs could capture the distribution of real sensor data of different activities.

D. Usability of Synthetic Sensor Data

The principal goal of SensoryGANs is to generate useful sensor data of human activity. The synthetic sensor data can be used to complement the small amount of raw data and to improve the classification accuracy of HAR. We use traditional machine learning methods to evaluate the quality of generated sensor data from the well-trained generators. The flow chart of evaluating the usability of synthetic sensor data from SensoryGANs is shown in Figure 6.

1) Data Preparation: After preprocessing these raw accelerometer data as consultant accelerometer data by Equation 4, we will get a group of raw consultant sensor data containing 194,436 stay samples, 204,046 walk samples and 200,691 jog samples. These three types of real sensor data will be used to train the corresponding SensoryGANs models respectively by Algorithm 1. After achieving the stability of training SensoryGANs models, the trained generators will be used to produce

TABLE II: Classification accuracy on real vs synthetic data

	SVM	Decision Tree	KNN	Logistic Regression	AdaBoost	Random Forest
Real Data	0.85714	0.79592	0.87714	0.87755	0.89796	0.87755
Synthetic Data	1.0	1.0	1.0	1.0	1.0	1.0
Hybrid Data	0.97938	0.97938	0.93814	0.95876	0.95876	0.97938

synthetic sensor data according to specified requirements. To avoid imbalance problem, we use trained generators to produce 200,000 synthetic sensor data for each activity. That is, we prepare synthetic data with the similar quantity of real data. And we set three group of comparative experiments. The first group contains all real sensor data. The second group contains all synthetic sensor data from the well-trained generators in SensoryGANs models. The third group contains both all the real sensor data and the synthetic sensor data.

2) Feature Extraction: Before classifying human activities, hand-engineered 27-dimension statistical features are extracted by sliding window strategy from these raw and synthetic consultant data. The size of sliding window is set as 50, and the length of time step is 25. We extract 27 features including the time domain features and the frequency features, such as mean, standard-deviation, max, min, mode, energy, peak-value, peak-position, shape-skew, shape-kurt, amplitude-skew. Time series sensor data in each sliding window are indicated as $X_t = \{x_1, x_2, ..., x_t\}$. x_t is the size of sliding windows. x is an item of consultant accelerometer data. Take the example of the mean of sensor data in one sliding window, we calculate it as equation 5.

$$mean = \frac{1}{t}\sum_{i=1}^{t} x_t \tag{5}$$

3) Classification and Analysis: After data preparation and feature extraction, all these feature vectors are then fed into classifiers in order to predict activity classes. We adapt six traditional classification algorithms as the basic classifier for evaluating the usability of SensoryGANs, such as SVM, Decision Tree, kNN, Logistic Regression, AdaBoost, Random-Forest. All the parameters for these classifiers are selected by grid search. As illustrated in Table II, the classification results of real data are desirable, and the classification accuracy of synthetic sensor data is perfect due to the optimal constructed and fine-tuned specific models for different activities. In fact, inaccurate segmentation for the raw sequential sensor data maybe leads to low classification accuracy compared with perfect results of synthetic data. Meanwhile, nonstandard actions and adventive noises in practical data collection process also result in data deviation from the true distribution. All these realities are inevitable and are the reasons that the real data are not as good as ideal synthetic data from SensoryGANs models. While the synthetic data generated by the well trained SensoryGANs models can present the implicit distribution of different activities.

Last but not the least, we use the hybrid sensor data containing both the real data and the synthetic data to verify

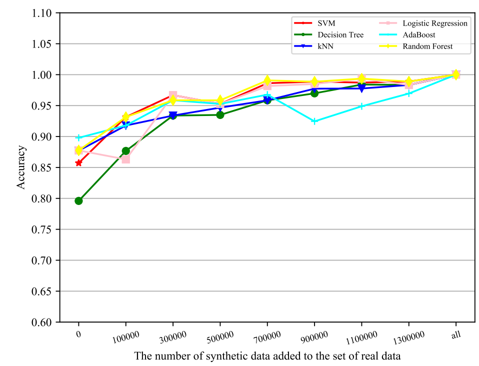


Fig. 7: Accuracy of classifcation with increasing number of synthetic data

if SensoryGANs models can effectively capture the implicit distribution of original sensor data of human activity. Results from the last row of Table II show that synthetic sensor data from SensoryGANs models can mostly improve the quality of original sensor data. The implicit distribution of original sensor data can be excavated by well-designed generative adversarial network models. This will be very useful for sensor-based researches.

E. Further Experiments of Synthetic Data

Above experiments focus on the classification performances based on the same amount of the real sensor data and synthetic sensor data. This is a practical measure to avoid the imbalance. In practics, we often supply the small real sensor dataset with different amounts of synthetic data to evaluate the robustness of SensoryGANs models when facing the imbalance problem. The same six typical traditional machine learning methods are also adapted to evaluate the performance of using synthetic data supplement. We used the well-trained generators from SensoryGANs models to generate seven synthetic datasets with different amounts, 100000, 300000, 500000, 700000, 900000, 1100000, 1300000 respectively.

Results are illustrated in Figure 7. The tick 0 on the x-axis represents that there are just real sensor data. The tick 'all' on the x-axis denotes that there are full synthetic sensor data which have the similar amount of real sensor data. Overall, with the increasing number of synthetic sensor data, we will observe better and better classification results. In detail, we can find that the performance of most classifiers get a little bit worse when 500000 synthetic sensor data are added into the real sensor dataset. There are 2.5 times synthetic sensor data than the real sensor data under this condition. This is a not supposed strange phenomenon which may be conquered by multiple tests. But the problem persists when we repeated the same experiments several times. Maybe this problem will be overcome with the improvement of our SensoryGANs.

VI. Conclusion and Future work

In the paper, we explore unbroken GANs framework to generate human activity sensor data for improving the performance of HAR, for the first time. Because Our SensoryGANs framework is hugely preliminary, we design specifical GANs models for three human daily activities respectively. At the forefront of an innovative research field, we also propose three visual evaluation methods for assessing the performance of SensoryGANs, the Local Visual Evaluation, the Global Visual Evaluation and the Memory Independence Visual Evaluation. We also evaluate the usability of synthetic sensor data to improve the performanc of human activity recognition. With the improvement of SensoryGANs, the research of human activity recognition especially in resource-constrained environments will be greatly encouraged.

In the future, we will explore a unified GANs framework to capture the implicit distributions of diverse human activities. Based on that, we will try to construct a creative GANs model that could produce new kinds of human activities which are unknown to current model at the training stage. At the same time, we will also consider the relationships of the 3-axis in motion sensors, which may be beneficial to improve learning ability of SensoryGANs.

Acknowledgment

This work is supported by the Natural Science Foundation of China under Grant No. 61572471

This work is supported in part by Natural Science Foundation of China under Grant No.61472399

References

[1] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," *Computers in Human Behavior*, vol. 15, no. 5, pp. 571–583, 1999.

[2] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008.

[3] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. d. R. Millán, and D. Roggen, "The opportunity challenge: A benchmark database for on-body sensor-based activity recognition," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 2033–2042, 2013.

[4] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors." *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[5] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.

[6] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, vol. 81, pp. 307–313, 2018.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[8] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5485–5493.

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arXiv preprint arXiv:1611.07004*, 2016.

[10] C. Ledig, Z. Wang, W. Shi, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, and A. Tejani, "Photo-realistic single image super-resolution using a generative adversarial network," 2016.

[11] F. Zhao, J. Feng, J. Zhao, W. Yang, and S. Yan, "Robust lstm-autoencoders for face de-occlusion in the wild," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 778–790, 2018.

[12] O. Press, A. Bar, B. Bogin, J. Berant, and L. Wolf, "Language generation with recurrent generative adversarial networks without pre-training," *arXiv preprint arXiv:1706.01399*, 2017.

[13] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient." in *AAAI*, 2017, pp. 2852–2858.

[14] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," in *Wearable and Implantable Body Sensor Networks (BSN), 2016 IEEE 13th International Conference on*. IEEE, 2016, pp. 71–76.

[15] T. Plötz, N. Y. Hammerla, and P. Olivier, "Feature learning for activity recognition in ubiquitous computing," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, no. 1, 2011, p. 1729.

[16] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[17] E. L. Denton, S. Chintala, R. Fergus *et al.*, "Deep generative image models using a laplacian pyramid of adversarial networks," in *Advances in neural information processing systems*, 2015, pp. 1486–1494.

[18] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," *arXiv preprint arXiv:1610.09585*, 2016.

[19] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[20] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," 2017.

[21] O. Mogren, "C-rnn-gan: Continuous recurrent neural networks with adversarial training," *arXiv preprint arXiv:1611.09904*, 2016.

[22] 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.

[23] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," in *Proceedings of the 6th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011, pp. 92–98.

[24] Y.-L. Hsu, S.-L. Lin, P.-H. Chou, H.-C. Lai, H.-C. Chang, and S.-C. Yang, "Application of nonparametric weighted feature extraction for an inertial-signal-based human activity recognition system," in *Applied System Innovation (ICASI), 2017 International Conference on*. IEEE, 2017, pp. 1718–1720.

[25] C. Chen, R. Jafari, and N. Kehtarnavaz, "Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 168–172.

[26] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, and R. Dürichen, "Cnn-based sensor fusion techniques for multimodal human activity recognition," in *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, 2017, pp. 158–165.

[27] M. Alzantot, S. Chakraborty, and M. B. Srivastava, "Sensegen: A deep learning architecture for synthetic sensor data generation," pp. 188–193, 2017.

[28] J. T. Springenberg, "Unsupervised and semi-supervised learning with categorical generative adversarial networks," *arXiv preprint arXiv:1511.06390*, 2015.

[29] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, 2016, pp. 2234–2242.

[30] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," 1999.