A Novel Deep Neural Network Method for HAR-Based Team Training Using Body-Worn Inertial Sensors

Yun-Chieh Fan 1,2, Yu-Hsuan Tseng 3 and Chih-Yu Wen 2,4,*

1 Simulator Systems Section, Aeronautical System Research Division, National Chung-Shan Institute of Science and Technology, Taichung 407, Taiwan
2 Department of Electrical Engineering, National Chung Hsing University, Taichung 402, Taiwan
3 Department of Computer Science and Engineering, National Chung Hsing University, Taichung 402, Taiwan
4 Innovation and Development Center of Sustainable Agriculture (IDCSA), National Chung Hsing University, Taichung 402, Taiwan
* Correspondence: cwen@dragon.nchu.edu.tw; Tel.: +886-04-2285-1549

Abstract: Human activity recognition (HAR) became a challenging issue in recent years. In this paper, we propose a novel approach to tackle indistinguishable activity recognition based on human wearable sensors. Generally speaking, vision-based solutions struggle with low illumination environments and partial occlusion problems. In contrast, wearable inertial sensors can tackle this problem and avoid revealing personal privacy. We address the issue by building a multistage deep neural network framework that interprets accelerometer, gyroscope, and magnetometer data that provide useful information of human activities. Initially, the stage of variational autoencoders (VAE) can extract the crucial information from raw data of inertial measurement units (IMUs). Furthermore, the stage of generative adversarial networks (GANs) can generate more realistic human activities. Finally, the transfer learning method is applied to enhance the performance of the target domain, which builds a robust and effective model to recognize human activities.

Keywords: human activity recognition; variational autoencoder; generative adversarial networks

1. Introduction

Body area network (BAN) [1] is a wireless network consisting of inertial measurement units (IMUs) that are placed over the body. The function of a BAN is to track the body movement and gestures using multiple sensors. The on-body inertial sensors are widely used in the domain of human activity recognition such as video gaming, medical application, image graphics, and skill training. In healthcare, sensor data can be utilized to monitor the daily activities of patients and detect sudden events such as falling down. In the military, due to the limited physical space and training facilities, body tracking can be utilized to train soldiers immersed in the battle field scenarios such as virtual training simulators, which can capture body movements such as walking, running, shooting, and crouching in a virtual environment. Considering the difficulty of tracking and recognizing locomotive motions from soldiers by using a single camera, the authors [2–4] exploit Microsoft Kinect to capture color and depth data from the human body. However, to recognize different human actions by tracking the front view of the human skeleton is a challenging issue. Therefore, the above methods may be not feasible in a real-time virtual environment for military training.

In addition, on-body sensors have merits regarding privacy concerns and do not reveal more personal privacy than cameras. Furthermore, inertial sensors can be worn anywhere regardless of fixed-cameras location for image recognition. Moreover, RGB cameras usually work in a sufficient lighting-source room. Therefore, in order to develop a training simulator that is capable of capturing soldiers’ accurate postures and movements, a BAN is the proper bridge to connect the physical and virtual environments. In addition, a
human activity recognition system can be built by wearable sensors on the human body, due to the advancements in lowering the energy consumption and increasing the computational power of inertial sensors.

In our previous work [5], a motion tracking system is proposed to monitor posture and gesture interactions between soldiers and the virtual environment via an inertial BAN method. The wireless BAN deployed over the full body of a soldier is composed of accelerometers, gyroscopes, and magnetometers. A kinematic chain is built to describe the relationship between rigid body movements and motions of joints. However, the main problem of wearable sensors is drifting in the wild. The wearable sensors deployed on the body keep changing positions and suffer from signal interference when the user is moving or performing activities. Moreover, the complementary filter in this motion-tracking method is ineffective to recognize a variety of rapid or complex actions. Paulich [6] presents evidence that the best error performance of a single wireless inertial sensor ranges from 0.51° to 1.65° RMS during human daily activity. In contrast, Marcard [7] indicates that a set of IMUs deployed over the body achieves only 16.9° average geodesic distance between ground-truth and IMUs, compared with a better result of 12.1° using vision and inertial fusion solution.

To address the above issues, we propose a multistage deep neural network (DNN), which is less dependent on the handcrafts feature extraction, comparing to other famous machine learning algorithms, such as support vector machine (SVM), naive Bayes, and random forest (RF) models. Since the amount of labelled data is not sufficient to train the DNN model on indistinguishable human activities, we leverage the issue and propose a novel approach to train unlabeled data, integrating sensor fusion, an autoencoder, a generative adversarial network (GAN), and a convolutional neural network (CNN) classifier. An autoencoder is mainly utilized for exploratory data analysis to discover patterns among data, is used in the pre-training process, and is efficient on feature extraction and dimensionality reduction. The key point for applying a GAN is to generate synthetic data for HAR and we adopt the transfer learning method to recognize activities of team members. We retrained the team modal from a well-trained GAN model via the limited amount of team training data. In addition, a dense layer is added on the top of classifier layers and the model parameters of the collective sensor data of several trainees are reused to retrain the team activity for advanced recognition of team activities. The transfer learning method helps enhance the performance of specific activity recognition.

Therefore, with the customized deep neural network method, the classification accuracy of the real dataset can be boosted. The proposed method helps avoid mismatched actions from the former single trainee activity classifier, which corrects the team training problems. The presented approach can achieve accuracy improvement compared to other methods [8–12]. The major contributions and key features of this paper are:

- Proposing a human activity recognition approach based on DNN for classification with a GAN model;
- Presenting a variational autoencoder (VAE) model method to denoise and extract the signal from inertial sensors in the environment;
- Using synthetic skeleton information to recognize the activities of individuals without using a vision approach;
- The proposed system can be scaled up to a ten-sensor deployment of the whole body or scaled down to two-sensor deployment of the human body;
- Recognizing the activities of a group or a team via pre-training model-based transfer learning. When it comes to the training dataset, the proposed method is limited due to insufficient activity data for training the neural network, especially when the transfer learning method cannot play a role in improving the accuracy of team activity recognition. With the help of a transfer learning approach, the model is simply retrained with the target data, conducted with fewer training epochs. Therefore, with scarce data, the proposed system can be utilized to achieve better classification accuracy.
The organization of this paper is as follows: Section 2 reviews related works about inertial-sensor-based HAR approaches. Section 3 presents the proposed system architecture. Section 4 evaluates the system performance and presents the performance comparison. Finally, Section 5 draws conclusions and shows future research directions.

2. Related Works

Inertial sensors have been utilized for human activity recognition due to several advantages, such as being wearable, light weight, and the ability to use plenty of sensors, where the IMUs can collect activity data over a period of time (e.g., detecting emergency events or recognizing daily activities). To classify different actions or activities, several machine learning approaches are proposed for feature extraction. For instance, k-nearest neighbor (kNN), SVM, RF [13], hidden Markov model (HMM) [14,15], and decision tree (DT) models. The authors [16–18] use machine learning such as kNN and SVM to detect the patient falling events. The authors of [19] and [20] present that the methods using SVMs are capable of improving the recognition accuracies. Considering the HAR as a general machine learning problem that requires feature extraction and feature selection, Suto et al. [21,22] emphasize the importance of feature selection on HAR problems, and further investigate the performance of feature selection methods in combination with machine learning algorithms, such as feed-forward artificial neural network, kNN, and decision tree. Notice that these methods have a high classification rate with respect to a few daily activities, and hand-crafted features are the key points for increasing their accuracies.

In recent years, artificial neural networks have shown a different impact on human activity recognition. The multilayer perceptron with multilayer feed-forward architecture can deal with complex activity classification for the ability of non-linear activation. DNN methods for action recognition were proposed in recent years [23,24]. Moreover, a CNN model can benefit the body skeleton and extract the information from the characterization of the skeleton of human, e.g., [25–27]. The authors in [26] use the CNN approach to acquire time-series data and build a multilayer classifier for sequences of skeleton recognition. Recurrent neural networks (RNNs) are also an approach to capture the spatial temporal evolution of the skeleton. The existing methods adopted gated RNN architectures such as the long short-term memory network (LSTM) [28,29]. The authors use the LSTM method to combine the video and the inertial sensors for pose prediction [30]. The work illustrates how a complex RNN model copes with the inertial sensors and a sequence of images for action classification [31]. The interested reader may refer to [32,33] for further discussion of action recognition using data sensing techniques and deep learning methods.

One of the crucial factors of raising the accuracy of human activity recognition rate is the sufficient amount of training data. However, acquiring sufficient data is difficult under this circumstance. Thus, a GAN is developed to learn a hidden structure of the data from the original distribution and then generate new data within the same distribution. Several GAN models [34–39] are proposed to generate verisimilar sensor data for recognizing the actions, where the generated data allow the use of data augmentation in HAR to improve the accuracy. The comparison of the state-of-the-art methods using sensory data with the GAN model are listed in Table 1.

To characterize the diversity with a multi-dimensional dataset, smart metasurfaces, platforms for multiple sensors, are proposed. As described in [43], the smart metasurface is capable of sensing ambient environments by integrating an additional sensor(s) and can adaptively adjust its electromagnetic operational functionality through an unmanned sensing feedback system. A motion-sensitive smart metasurface integrated with a three-axis gyroscope is developed to illustrate the feasibility of the metasurface platform. Moreover, based on a multi-layer digital-coding metasurface array, Liu et al. [44] report a programmable diffractive deep neural network for handling various applications such as wireless communications, signal enhancement, medical imaging, remote control, and the Internet of Things, which may further extend the applications of the proposed algorithm framework.
Table 1. The comparison of different sensory methods using for HAR with GAN model.

| Methods          | Purpose | IMUs    | Advantages                                                                 | Limitations                                                                 |
|------------------|---------|---------|----------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| SensoryGAN [38]  | Augmentation | Acc     | The first work about data augmentation with synthesizing sensory data.     | Using three activity-specific GANs for three activities.                      |
| Zhang [40]       | Augmentation | Acc Gyro Mag | The first method that uses semi-supervised GAN for activity recognition. | Using some hand-crafted parameters to control the classification results.      |
| Soleimani [41]   | Augmentation | Acc     | Generating data of the target domain from the source domain via GANs.     | The diversity of the sensory data is limited.                               |
| Mathur [42]      | Augmentation | Acc Gyro | Enriching the training set with sufficient discrepancy of different sensor deployments. | The diversity of the synthetic data is limited and not guaranteed.          |
| Proposed Method  | Augmentation | Acc Gyro Mag | Using synthetic sensory data to recognize the activities of individuals and a group. | Using limited team data to train the neural network.                      |

3. System Description

As mentioned above, human action recognition is crucial to develop an immersive training simulator. The wireless BAN can immediately track the action of the skeleton of the trainee. The system operations consist of four stages: (1) sensor fusion, (2) a VAE, (3) a GAN, and (4) CNN-based classification. We use a VAE to pre-train the sensing data and extract crucial features. A VAE encodes the nine-axis data into a feature-representation vector, constrained by Gaussian distributions. The VAE stage aims to denoise the sensing measurements, which transforms into a family of Gaussian distributions in the latent space with mean and variance. Our goal is to build 2D skeleton maps from the inertial sensors and to recognize the activities from the skeleton maps. To deal with the data from the VAE process, the deep convolutional GAN (DCGAN), which consists of two major modules, a skeleton map generation module and an activities recognition module, is applied. Figure 1 shows the proposed DNN architecture for human activity recognition.

![Figure 1. The proposed DNN architecture on HAR.](image)

3.1. Stage 1: Sensors Fusion

This section introduces the sensors used in the BAN. The sensor nodes are nine axial inertial measurement units (IMUs), which are equipped with an ARM Cortex-M4 microprocessor and ST-MEMS chips. Figure 2 illustrates the appearance of the sensor hardware. The customized sensor can capture three-axis acceleration, three-axis angular...
velocity, and three-axis magnetic strength, which transmits data streaming via 2.4 GHz RF radio. The trainees wear multiple sensors on the arms, wrists, thighs, lower legs, and back. As depicted in Figure 3, the inertial signals are acquired from accelerometers, gyroscopes, and magnetometers for the T-pose action. The orientation of the sensors that are attached to human bones are measured in local frames employed in quaternion representation.

\[
q_{t_1}^{\text{node}} = q_{t_0}^{\text{node}} \odot q_{t_1}^{\text{node}},
\]

where \(\odot\) is a quaternion multiplication operator, \(q_{t_0}^{\text{node}}\) is defined as a sensor node’s quaternion output while the wearable sensor nodes is in the pose at time \(t_0\), and \(q_{t_1}^{\text{node}}\) depicts a valid rotation after a time interval \(\Delta t\). Note that \(t_1 = t_0 + \Delta t\) and \(q_{\Delta t}^{\text{node}}\) is the rotation used for the orientation of the sensor node.

![Figure 2. Top view of sensor nodes (left); the wearable sensor node (right).](image)

**Figure 2.** Top view of sensor nodes (left); the wearable sensor node (right).

![Figure 3. The deployment of inertial sensors over the body.](image)

**Figure 3.** The deployment of inertial sensors over the body.

3.2. *Stage 2: Variational Autoencoder (VAE)*

A VAE [45,46] can preserve the source information from sensor nodes, including three-axis acceleration, three-axis angular velocity, and three-axis magnetic strength data. The VAE aims to handle the sensor data from each IMU and transform better information from the orientation of sensors with measured noise. A sequence of sensors \(M = [s_1, s_2, ..., s_9]\) is given from upper stage and becomes the input of the VAE, which consists of an encoder and a decoder. Figure 4 depicts the proposed VAE architecture, which contains both a CNN-based encoder part and decoder part for the data pre-training process. Both of them have three convolutional layers. The multiple layers of the VAE denoise data and extract the source information. In the encoding process, the input data are transformed into features that can reconstruct valuable inputs from the source. The decoding process is conducted in a similar way to perform influences after the three convolutional layers.
We used 100 batches to train the VAE model and the Adam optimizer to minimize the loss with a learning rate of 0.01. The VAE reconstructed the sensing data as a less noisy signal and preserved the crucial information of inertial sensors. To extract the useful information from IMU sensors, three convolutional hidden layers and three deconvolutional layers were used for the encoder and the decoder. The loss function is defined as follows:

\[ L(x) = E_{q(z|x)} \log p(x \mid z) - D_{KL}[q(z|x) \parallel p(z)], \]  

where \( q(z \mid x) \) is the output of the convolutional encoder part, \( D_{KL} \) is the Kullback–Leibler distance between \( p \) and \( q \), and \( p(z) \) is the prior information for the latent variable and unit-variance Gaussian,

\[ p(x \mid z) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x - f(z))^2}{2\sigma^2}}, \]

where \( \sigma^2 \) is the variance and \( p(x \mid z) \) is the output of the convolutional decoder.

Figure 4. The proposed VAE architecture for data pre-training.

3.3. Stage 3: Generative Adversarial Network (GAN)

Referring to the operations of the first two stages, a GAN is capable of producing a sequence of synthetic skeletons, consisting of a generator that attempts to capture the distribution and a discriminator that assesses the generator output and penalizes unrealistic samples. It is a network architecture that is trained to achieve an equilibrium. We receive sensor data from VAEs and then feed the data into GANs in order to generate entirely new and realistic skeleton maps that match the distribution of a given target dataset. The architecture for the proposed GAN model is illustrated in Figure 5. The generator yields a sequence of skeleton as the input of the discriminator. Uniquely, our approach utilizes a conditional GAN to generate a sequence of verisimilar skeletons from individuals on different activities. A small amount of sensor data of individuals are trained as labeled data with respect to the human activities. Given multiple sensor data of an individual, the generator tries to create a corresponding skeleton map for data augmentation. We train a sequence of the skeletons so that inferred skeletons are similar to real ones. Thus, the skeleton map derived from the generator is applied to distinguish the human activities by a discriminator.
The two modules, a sequence of skeleton generator and an activities discriminator, are trained by using the G and D loss functions, respectively. The GAN training is a strategy to define a game between the generator network and the discriminator network. The block diagrams of the generator and discriminator are illustrated in Figures 6 and 7.

\[ L = \max_D \min_G V(D, G) = \min_G \max_D E_{x \sim P_{data}(x)} [\log (D(x))] + E_{z \sim P_z(z)} [\log(1 - D(\hat{x}))], \]  

(4)

where \( L \) is the sum of GAN stage loss, \( V \) is the value of this minimax game strategy, \( z \) is a vector of random values, \( P_{data} \) is the prior probability distribution of the real data, \( P_z \) is the prior probability distribution of random input vectors \( z \), and \( x \) is a sequence of the input VAE data. Since a GAN model is difficult to train and optimize from the generator’s output rather than the discriminator’s, a Wasserstein GAN (WGAN) [47] is used for IMUs data prediction. In the WGAN, we now utilize a gradient penalty to optimize the generator process.

\[ L_G = L_{gan} + \alpha L_p, \]  

(5)

\[ L_{gan} = -D(\hat{x}) = G(z), \]  

(6)

\[ L_p = \left\| \nabla \hat{x} D(\hat{x}) \right\|^2, \]  

(7)

where \( L_G \) is the generator loss function, \( L_{gan} \) is the original WGAN critic loss, and \( L_p \) is the gradient penalty loss, where the operator \( \| \| \) indicates concatenation.

![Figure 5. The proposed generative adversarial network.](image)

![Figure 6. The block diagram of the generator.](image)
3.4. Stage 4: CNN Classifier

A CNN architecture is utilized to perform the classification of skeleton sequences derived from the output of Stage 3. The design of the CNN classifier is illustrated in Figure 8. Observe that the CNN consists of three convolutional layers, including a batch normalization layer, a max pooling layer, and a dense layer. Afterwards, the classification is performed via a SoftMax function, which yields:

$$\text{Softmax classifier} = P(c|p) = \frac{\exp(p^{L-1}w_c^L + b^L)}{\sum_k^{NC}(p^{L-1}w_k^L)},$$  

where $c$ is the activities class, $L$ is the last layer index, $N_c$ is the total number of activities class, $w$ is the weight, and $b$ is the bias. We pre-train the CNN classification using augment data generated from GANs and fine-tune the model by using the real data.

3.5. Stage 5: Transfer Learning

Since collecting a sufficient amount of team training data may be difficult, there is a need for transfer learning methods from a well-trained source domain to the target team.

**Figure 7.** The block diagram of the discriminator.

**Figure 8.** The design of the CNN classifier.
training domain. Hence, we can use the GAN model to train the little labeled data on mutual actions. The model with transfer learning is as efficient as the modal in the source domain. Moreover, a hidden layer is added on the top of the CNN classifier layers for classification [48,49]. The transfer learning method is applied to enhance the performance of the target domain, which is lacking a sufficient amount of the labeled data. We kept the optimizer, batch size, and learning rate the same as the previous model and retrained the current model with transfer learning, which prevented forgetting the previously learned knowledge for obtaining better classification accuracy in the target domain.

4. System Evaluation

In order to evaluate the proposed system, multiple experiments are conducted for human activities recognition on the public dataset: UCI OPPORTUNITY dataset. The software is Tensorflow Keras for Python. The computer hardware is equipped with 3.1 GHz Intel Core i7 processor, 8GB RAM, and a NVidia 2070 graphic card.

4.1. Data Feature

The UCI dataset is applied, which features four subjects who perform 17 different activities, such as opening a door, opening a drawer, and toggling a switch. The data are recorded every 5 min by using seven inertial sensors that are attached to the torso, right upper/lower arm, left upper/lower arm, right leg, and left leg. The sensor data of different activities recorded at a frame rate of 32 Hz were segmented into multiple 2 s windows.

4.2. Data Processing

The data from nine different measurements of the sensors (x, y, z gyroscopes, x, y, z accelerometers, x, y, z magnetometers) are preprocessed, normalized, and transferred to quaternion forms. Therefore, the pre-trained data are prepared for the variational autoencoder stage for training. The output of VAE is applied as the input for the GAN stage. After the GAN training, the generated synthetic output is processed via the final stage (i.e., a CNN classifier). The generator and discriminator are trained with the Adam optimizer with a learning rate of $1 \times 10^{-5}$. A batch size of 100 is used for the GAN architecture, where the generator loss and discriminator loss converge to 59.92 and 27.68 after 100 training batches, respectively. The best generative loss is 4.41 for epoch 91. The result of generator and discriminator losses are depicted in Figure 9.

**Figure 9.** The discriminator and generator losses with a total of 100 epochs during the training process (blue: discriminator loss. Orange: generator loss).
4.3. Assessing the Generated Data

We compare the generated data to several machine learning methods such as SVM, kNN, and decision tree. All parameters used in the machine learning methods are grid search. As shown in Table 2, the proposed method outperforms other machine learning methods in terms of average accuracies, which suggests that more information from the IMUs in our proposed method leads to better results, compared with other machine learning methods. There is an increment in the performance metrics for three different methods if the amount of IMUs increase, except in the decision tree method. The average accuracy of the SVM method increases from 90.6% to 92.9%, the kNN method from 89.1% to 90.6%, and the proposed method from 91.5% to 95.7%. Note that different activities are usually classified for upper or lower body movements. Hence, if we place fewer IMUs over the body, the accuracy is highly related to the sensor deployment strategy, which can obviously influence the accuracy of the result. Accordingly, in addition to using two wearable IMUs over the human body, we also present a multi-IMU deployment method that can be applied to clearly describe body movements. The DNN architecture is built to tackle specific HAR cases.

|        | One IMU (%) | Two IMUs (%) | Five IMUs (%) | Average Result (%) |
|--------|-------------|--------------|---------------|--------------------|
| SVM    | 90.6        | 90.8         | 92.9          | 91.4               |
| Decision tree | 91.4 | 87.1         | 76.2          | 84.4               |
| kNN    | 89.1        | 89.5         | 90.6          | 89.7               |
| Proposed method | 91.5 | 92.1         | 95.7          | 93.1               |

Moreover, to assess the impact of a random noise in the environment on system performance, we examine the performances of the proposed method, DNN method with CNN model, and LSTM model. As shown in Table 3, we summarize the quantitative results of mean accuracy and F1 score on sensory data with Gaussian noise (signal to noise ratio of 7.78 dB). In the training process, all the models were trained by feeding Gaussian noise to a sensor of the left leg in the experiment. Referring to the results, the proposed method outperforms other DNN models on HAR if there is a random noise in the environment. The mean accuracy reaches 88.4%, which is superior to the RNN (LSTM) model in [50] with 82.9% accuracy for the same selection of noisy sensor on test data of every subject.

| Metric          | Model | CNN Model [50] | RNN (LSTM) Model [50] | Proposed Method |
|-----------------|-------|----------------|-----------------------|-----------------|
| Accuracy (without noise) | 94.5% | 95.4%          | 95.9%                 |
| F1 score (without noise) | 93.7% | 95.4%          | 95.8%                 |
| Accuracy (with noise) | 78.5% | 82.9%          | 88.4%                 |
| F1 score (with noise) | 77.4% | 81.7%          | 86.9%                 |

4.4. Classification Performance

To validate the classification performance, we evaluate the confusion matrix result on the UCI OPPORTUNITY dataset. As depicted in Figure 10, the accuracy of all activities trained by all sensors data ranges from 86% to 97%. Although the accuracy of open drawer, close drawer cases have lower classification results, the proposed method is able to benefit...
from the DNN model to improve the accuracy of most classification results. We also evaluate mutual action classification results.

| (N) | Open Door | Open Door | Close Door | Close Door | Open Fridge | Open Fridge | Open Dish | Open Dish | Close Dish | Close Dish | Open Door2 | Open Door2 | Close Door2 | Close Door2 | Close Table | Close Table | Stand | Stand | Stand | Stand |
|-----|-----------|-----------|------------|------------|-------------|-------------|-----------|-----------|------------|------------|------------|------------|------------|------------|------------|------------|--------|--------|--------|--------|
| 95  | 3         | 1         | 1          |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 4   | 96        | 3         |            |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 95  | 3         |            |            |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 1   | 96        | 3         |            |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 97  | 3         |            |            |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 1   | 92        |            |            |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 2   | 80        | 4         | 4          |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 1   | 3         | 19        | 19         | 3          | 3           |            |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 8   | 9         | 10        | 10         |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 2   | 10        | 4         | 4          |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 2   | 2         | 2         | 2          | 0          | 0           |            |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 3   | 1         | 1         | 1          | 4          | 4           |            |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 2   | 3         |            |            |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 3   |            |            |            |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 96  |            |            |            |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |
| 97  |            |            |            |            |             |             |           |           |            |            |            |            |            |            |            |        |        |        |        |

**Figure 10.** The confusion matrix result of our DNN method on different human activities.

The parameters are learned by the small amount of team training data. We combine two subjects with different actions (e.g., open a door, close a door) to be the target domain. The other actions can be the source domain. By using transfer learning method, we train the GAN model with target sensor data. Hence, the proposed model is able to recognize the different activities from small amounts of data. Accordingly, with the transfer learning method, the proposed method is trained on the UCI OPPORTUNITY dataset, which means the parameters of the hidden layer are renewed by a small amount of training data. Finally, the mean accuracy leads to an improvement of 12% compared with the proposed architecture, without introducing the transfer learning method.

Therefore, our solution raises the accuracy of motion recognition by building a DNN model on the measurement of sensors. The synthetic action sequences generated from the IMUs are deployed on five different subjects for two seconds. We depicted the data in a visual presentation. Figure 11 shows the sitting down action generated from the GAN model, which is able to enforce the motion sequence.

**Figure 11.** The synthetic sequences derived from the proposed DNN method, shown in 3D presentation (a sitting down action was performed by five different subjects).
5. Conclusions

In this work, we propose a novel and supervised method for human activity recognition. The VAE and GAN models are able to improve the accuracy of action classification. Accordingly, we can recognize the motion of IMUs data even though the inertial signal is drifting over a long duration, which opens another door for our simulator to execute a correct motion or pose in the virtual reality operation. The on-body sensor deployment of the proposed system can be scaled up (e.g., a ten-sensor deployment) or scaled down (e.g., a two-sensor deployment). Compared to the existing deep, convolutional, and recurrent models in [50], the proposed method leads to significant improvement in HAR accuracy. When it comes to the training dataset, the proposed method is limited due to insufficient activity data for training the neural network, especially when the transfer learning method cannot play a role in improving the accuracy of team activity recognition, as described in Section 3.5 (page 8). Ideally, higher classification accuracy can be achieved with a more complete sensor dataset in the training process. In future work, we plan to collect more on-body HAR data and conduct efficient sensor deployment strategies so that the classification accuracy can be further improved.

Author Contributions: Conceptualization, Y.-C.F. and C.-Y.W.; formal analysis, Y.-C.F. and Y.-H.T.; funding acquisition, C.-Y.W.; methodology, Y.-C.F., Y.-H.T. and C.-Y.W.; supervision, C.-Y.W.; visualization, Y.-C.F.; Writing—original draft, Y.-C.F.; writing—review and editing, C.-Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science and Technology of Taiwan under grant number MOST-110-2634-F-005-006, and by the “Innovation and Development Center of Sustainable Agriculture” from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The author at the Simulator Systems Section, Aeronautical System Research Division, National Chung-Shan Institute of Science and Technology, Taiwan, was the subject of the experiments. The author consented to participate in this research study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References
1. Ambroziak, S.J.; Correia, L.M.; Katulski, R.J.; Mackowiak, M.; Oliveira, C.; Sadowski, J.; Turbic, K. An Off-Body Channel Model for Body Area Networks in Indoor Environments. IEEE Trans. Antennas Propag. 2016, 64, 4022–4035. [CrossRef]
2. Papadopoulos, G.T.; Axenopoulos, A.; Daras, P. Real-Time Skeleton-Tracking-Based Human Action Recognition Using Kinect Data. In Proceedings of the MMM 2014, Dublin, Ireland, 6–10 January 2014.
3. Kwon, B.; Kim, J.; Lee, S. An enhanced multi-view human action recognition system for virtual training simulator. In Proceedings of the 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, Korea, 13–16 December 2016; IEEE: Manhattan, NY, USA, 2016.
4. Liu, T.; Song, Y.; Gu, Y.; Li, A. Human Action Recognition Based on Depth Images from Microsoft Kinect. In Proceedings of the 2013 Fourth Global Congress on Intelligent Systems, Hong Kong, China, 3–4 December 2013; IEEE: Manhattan, NY, USA, 2013; pp. 200–204. [CrossRef] [PubMed]
5. Fan, Y.-C.; Wen, C.-Y. A Virtual Reality Soldier Simulator with Body Area Networks for Team Training. Sensors 2019, 19, 451. [CrossRef] [PubMed]
6. Paulich, M.; Schepers, M.; Rudigkert, N.; Bellusci, G. Xsens Mtiw Awinda: Miniature Wireless Inertial-Magnetic Motion Tracker for Highly Accurate 3D Kinematic Applications; Xsens: Enschede, The Netherlands, 2018; pp. 1–9.
7. von Marcard, T.; Henschel, R.; Black, M.J.; Rosenhahn, B.; Pons-Moll, G. Recovering Accurate 3D Human Pose in the Wild Using IMUs and a Moving Camera. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 614–631. [CrossRef]
8. Lan, T.; Wang, Y.; Yang, W.; Robinovitch, S.N.; Mori, G. Discriminative Latent Models for Recognizing Contextual Group Activities. IEEE Trans. Pattern Anal. Mach. Intell. 2011, 34, 1549–1562. [CrossRef] [PubMed]
9. Ibrahim, M.S.; Muralidharan, S.; Deng, Z.; Vahdat, A.; Mori, G. A Hierarchical deEP Temporal Model for Group Activity Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Amsterdam, The Netherlands, 11–14 October 2016; pp. 1971–1980.

10. Weber, M.; Auch, M.; Doblander, C.; Mandl, P.; Jacobsen, H.-A. Transfer Learning with Time Series Data: A Systematic Mapping Study. *IEEE Access* 2021, 9, 165409–165432. [CrossRef]

11. Wang, J.; Chen, Y.; Hu, L.; Peng, X.; Philip, S. Stratified Transfer Learning for Cross-Domain Activity Recognition. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), Athens, Greece, 19–23 March 2018; IEEE: Manhattan, NY, USA, 2018; pp. 1–10.

12. Chen, C.; Miao, Y.; Lu, C.X.; Xie, L.; Blunsom, P.; Markham, A.; Trigoni, N. MotionTransformer: Transferring Neural Inertial Tracking between Domains. In Proceedings of the AAAI Conference on Artificial Intelligence, Atlanta, Georgia, 8–12 October 2019; Volume 33, pp. 8009–8016. [CrossRef]

13. Bedogni, L.; Di Felice, M.; Bononi, L. By Train or by Car? Detecting the User’s Motion Type through Smartphone Sensors Data. In Proceedings of the 2012 IFIP Wireless Days (WD), Dublin, Ireland, 21–23 November 2012; pp. 1–6.

14. Lester, J.; Choudhury, T.; Kern, N.; Borriello, G.; Hannaford, B. A Hybrid Discriminative/Generative Approach for Modeling Human Activities. In Proceedings of the IJCAI’05 19th International Joint Conference on Artificial intelligence, Edinburgh, UK, 30 July–5 August 2005; pp. 766–772.

15. Łukowicz, P.; Ward, J.A.; Junker, H.; Starner, T. Recognizing Workshop Activity Using Body Worn Microphones and Accelerometers. In *Pervasive Computing*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 18–23.

16. Kepski, M.; Kwolek, B. Fall detection using body-worn accelerometer and depth maps acquired by active camera. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2016; Volume 9648.

17. Kepski, M.; Kwolek, B. Detecting human falls with 3-axis accelerometer and depth sensor. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 770–773. [CrossRef]

18. Zhang, T.; Wang, J.; Xu, L.; Liu, P. Using Wearable Sensor and NMF Algorithm to Realize Ambulatory Fall Detection. In *Advances in Natural Computation*; Springer: Berlin, Germany, 2006; pp. 488–491.

19. Huyhn, T.; Schiele, B. Towards less supervision in activity recognition from wearable sensors. In Proceedings of the 2006 10th IEEE International Symposium on Wearable Computers, Montreux, Switzerland, 11–14 October 2006; pp. 3–10.

20. Krause, A.; Ihmig, M.; Rankin, E.; Leong, D.; Gupta, S.; Siewiorek, D.; Smajagic, A.; Deisher, M.; Sengupta, U. Trading off prediction accuracy and power consumption for context-aware wearable computing. In Proceedings of the Ninth IEEE International Symposium on Wearable Computers (ISWC’05), Osaka, Japan, 18–21 October 2005; pp. 20–26.

21. Suto, J.; Oniga, S.; Sitar, P.P. Feature Analysis to Human Activity Recognition. *Int. J. Comput. Commun. Control* 2016, 12, 116–130. [CrossRef]

22. Suto, J.; Oniga, S.; Sitar, P.P. Comparison of wrapper and filter feature selection algorithms on human activity recognition. In Proceedings of the 2016 6th international conference on computers communications and control (ICCCC), Oradea, Romania, 10–14 May 2016; pp. 124–129.

23. Delachaux, B.; Rebetez, J.; Perez-Uribe, A. Mejia HFS Indoor activity recognition by combining one-vs.-all neural network classifiers exploiting wearable and depth sensors. In *Proceedings of the International Work Conference on Artificial Neural Networks, Tenerife, Spain*, 12–14 June 2013; Springer: Berlin, Heidelberg, 2013; pp. 216–223.

24. Nweke, H.F.; Teh, Y.W.; Al-Garadi, M.A.; Alo, U.R. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Syst. Appl.* 2018, 105, 233–261. [CrossRef]

25. Tan, M.; Le, Q.V. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv* 2019, arXiv:1905.11946.

26. Yang, J.; Nguyen, M.N.; San, P.P.; Li, X.L.; Krishnaswamy, S. Deep convolutional neural networks on multichannel time series for human activity recognition. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.

27. Avilés-Cruz, C.; Ferreyra-Ramírez, A.; Zuñiga-López, A.; Villegas-Cortez, J. Coarse-Fine Convolutional Deep-Learning Strategy for Human Activity Recognition. *Sensors* 2019, 19, 1556. [CrossRef] [PubMed]

28. Liu, J.; Shahroudy, A.; Xu, D.; Wang, G. Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition. In *Computer Vision–ECCV 2016*; Springer: Cham, Switzerland, 2016; pp. 816–833. [CrossRef]

29. Alahi, A.; Goel, K.; Ramanathan, V.; Robicquet, A.; Fei-Fei, L.; Savarese, S. Social Istm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 961–971.

30. Trumble, M.; Gilbert, A.; Malleson, C.; Hilton, A.; Colomosse, J. Total Capture: 3D Human Pose Estimation Fusing Video and Inertial Sensors. In Proceedings of the British Machine Vision Conference (BMVC), London, UK, 4–7 September 2017.

31. Imran, J.; Raman, B. Evaluating fusion of RGB-D and inertial sensors for multimodal human action recognition. *J. Ambient Intell. Humaniz. Comput.* 2019, 11, 189–208. [CrossRef]

32. Shweta; Khandnor, P.; Kumar, N. A survey of activity recognition process using inertial sensors and smartphone sensors. In Proceedings of the 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 5–6 May 2017; pp. 607–612. [CrossRef]
33. Chen, K.; Zhang, D.; Yao, Y.; Guo, B.; Yu, Z.; Liu, Y. Deep learning for sensor-based human activity recognition: Overview, challenges and opportunities. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–40. [CrossRef]

34. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X. Metaxas DN StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1947–1962. [CrossRef] [PubMed]

35. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; pp. 2672–2680.

36. Alzantot, M.; Chakraborty, S.; Srivastava, M. SenseGen: A deep learning architecture for synthetic sensor data generation. In Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops, Big Island, HI, USA, 13–17 March 2017; pp. 188–193.

37. Norgaard, S.; Saeedi, R.; Sasani, K. Gebremedhin AH Synthetic Sensor Data Generation for Health Applications: A Supervised Deep Learning Approach. In Proceedings of the the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), Honolulu, HI, USA, 18–21 July 2018; pp. 1164–1167.

38. Wang, J.; Chen, Y.; Gu, Y.; Xiao, Y.; Pan, H. SensoryGANs: An Effective Generative Adversarial Framework for Sensor-based Human Activity Recognition. In Proceedings of the International Joint Conference on Neural Networks, Rio de Janeiro, Brazil, 8–13 July 2018.

39. Si, C.; Wang, W.; Wang, L.; Tan, T. Multistage adversarial losses for pose-based human image synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.

40. Zhang, X.; Yao, L.; Yuan, F. Adversarial variational embedding for robust semi-supervised learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorag, AK, USA, 4–8 August 2019; pp. 139–147.

41. Soleimani, E.; Nazerfard, E. Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. *Neurocomputing* **2020**, *426*, 26–34. [CrossRef]

42. Mathur, A.; Zhang, T.; Bhattacharya, S.; Velickovic, P.; Joffe, L.; Lane, N.D.; Kawsar, F.; Lio, P. Using Deep Data Augmentation Training to Address Software and Hardware Heterogeneities in Wearable and Smartphone Sensing Devices. In Proceedings of the 2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Porto, Portugal, 11–13 April 2018; IEEE: Manhattan, NY, USA, 2018; pp. 200–211.

43. Ma, Q.; Bai, G.D.; Jing, H.B.; Yang, C.; Li, L.; Cui, T.J. Smart metasurface with self-adaptively reprogrammable functions. *Light. Sci. Appl.* **2019**, *8*, 98. [CrossRef] [PubMed]

44. Liu, C.; Ma, Q.; Luo, Z.J.; Hong, Q.R.; Xiao, Q.; Zhang, H.C.; Miao, L.; Yu, W.M.; Cheng, Q.; Li, L.; et al. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nat. Electron.* **2022**, *5*, 113–122. [CrossRef]

45. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.

46. Esser, P.; Sutter, E.; Ommer, B. A variational u-net for conditional appearance and shape generation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 8–22 June 2018; pp. 8857–8866.

47. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. *NIPS’17*. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 4–9 December 2017; pp. 5769–5779.

48. Wen, L.; Gao, L.; Li, X. A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *49*, 136–144. [CrossRef]

49. Morales, F.J.O.; Roggen, D. Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers—ISWC’16*, Heidelberg, Germany, 12–16 September 2016; ACM: New York, NY, USA, 2016; pp. 92–99.

50. Hammerla, N.Y.; Halloran, S.; Plötz, T. Deep, convolutional, and recurrent models for human activity recognition using wearables. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, New York, NY, USA, 9–15 July 2016.