mmFit: Low-Effort Personalized Fitness Monitoring Using Millimeter Wave

Yucheng Xie*, Ruizhe Jiang*, Xiaonan Guo*, Yan Wang†, Jerry Cheng‡, Yingying Chen§
*Indiana University-Purdue University Indianapolis, IN 46202, USA
†Temple University, Philadelphia, PA 19122, USA
‡New York Institute of Technology, New York, NY 10023, USA
§Rutgers University, New Brunswick, NJ 08901, USA
Email: *{yx11, ruizjian, xg6}@iupui.edu
,†y.wang@temple.edu,‡jcheng18@nyit.edu,§yingche@scarletmail.rutgers.edu

***Abstract*—There is a growing trend for people to perform workouts at home due to the global pandemic of COVID-19 and the stay-at-home policy of many countries. Since a self-designed fitness plan often lacks professional guidance to achieve ideal outcomes, it is important to have an in-home fitness monitoring system that can track the exercise process of users. Traditional camera-based fitness monitoring may raise serious privacy concerns, while sensor-based methods require users to wear dedicated devices. Recently, researchers propose to utilize RF signals to enable non-intrusive fitness monitoring, but these approaches all require huge training efforts from users to achieve a satisfactory performance, especially when the system is used by multiple users (e.g., family members). In this work, we design and implement a fitness monitoring system using a single COTS mmWave device. The proposed system integrates workout recognition, user identification, multi-user monitoring, and training effort reduction modules and makes them work together in a single system. In particular, we develop a domain adaptation framework to reduce the amount of training data collected from different domains via mitigating impacts caused by domain characteristics embedded in mmWave signals. We also develop a GAN-assisted method to achieve better user identification and workout recognition when only limited training data from the same domain is available. We propose a unique spatial-temporal heatmap feature to achieve personalized workout recognition and develop a clustering-based method for concurrent workout monitoring. Extensive experiments with 14 typical workouts involving 11 participants demonstrate that our system can achieve 97% average workout recognition accuracy and 91% user identification accuracy.**

***Index Terms*—mmWave sensing, fitness monitoring, generative adversarial network, domain adaptation training**

I. Introduction

Due to the global pandemic of COVID-19, we have been confined at home for a long period of time. The consequent physical inactivity might cause health problems, such as heart diseases, diabetes, and cancers [1]. As a result, many people perform physical exercises regularly at home for convenience, safety, and flexibility. Though there are various at-home workout and fitness programs available to help users achieve certain fitness goals, the lack of professional guidance and monitoring can make this process less effective. For example, you might not be able to do the exercises in correct positions and forms; your workout is not consistent or imbalanced between aerobic and strength training. Therefore, it is important to have an in-home fitness monitoring system that can track the exercise process of users and give them useful suggestions.

Traditional camera-based fitness monitoring may raise serious privacy concerns [2, 3], while sensor-based methods require users to wear dedicated devices [4, 5]. Some researchers propose to utilize WiFi signals for non-intrusive fitness monitoring [6, 7], but WiFi signals are sensitive to interference and surrounding environment changes. In recent years, millimeter wave (mmWave) signals have emerged for activity recognition since they enable higher-resolution sensing given their short wavelengths and high bandwidths. In addition, mmWave has already been integrated into the next-generation WiFi standards (i.e., IEEE 802.11ad). For example, human skeleton systems are proposed to capture human postures [8, 9, 10]. However, these systems require extra cameras (e.g., Microsoft Kinect) to provide accurate joint locations. Moreover, some activity recognition systems have shown satisfying performance for one-person scenarios [11, 12, 13]. Since, in reality, multiple persons may be present in a shared space, a system suitable for concurrent workout scenarios is desired for fitness monitoring.

Although mmWave has the potential to offer higher signal resolution comparing to traditional RF-based approaches, mmWave-based solutions still face technical challenges and obstacles for real deployment. First, existing mmWave-based methods need to collect enough amount of data for training machine learning models up to a satisfactory level of accuracy [14, 9, 12]. For example, they require users to repeat the activity moves many times. This makes it inconvenient and time-consuming for practical usage. Second, the trained machine learning models are environment-specific and re-training processes are required before they can be applied to a new environment [8, 15, 16]. Third, even in the same environment, users can work out in different spots. These small changes in locations can have a significant impact on the system performance [17, 11]. In summary, it is essential to develop a multi-person fitness monitoring system robust of environmental variations with light efforts in training data collection.

Toward this end, we propose *mmFit*, a mmWave-based personalized fitness monitoring system using only a single COTS mmWave device. The *mmFit* system integrates modules of workout recognition, user identification, multi-user monitoring, and training effort reduction. Specifically, our system records mmWave signals reflected from human bodies and captures

This is the author's manuscript of the article published in final edited form as:

Y. Xie, R. Jiang, X. Guo, Y. Wang, J. Cheng and Y. Chen, "mmFit: Low-Effort Personalized Fitness Monitoring Using Millimeter Wave," 2022 International Conference on Computer Communications and Networks (ICCCN), Honolulu, HI, USA, 2022, pp. 1-10, doi: 10.1109/ICCCN54977.2022.9868878

fine-grained workout information, such as workout types, repetitions, and participants. For workout recognition and user identification, a new spatial-temporal heatmap is proposed to capture workout dynamics using various activity characteristics (i.e., velocity, range of movements, and time duration). For multi-person monitoring, we utilize a clustering-based method to construct a spatial-temporal heatmap for each user based on the reflected signals. We then develop deep neural network models to extract unique high-dimensional features from the heatmaps to perform the workout and user identification tasks.

In addition, we develop a generative adversarial network (GAN) to synthesize a large number of virtual workout segments based on a small number of segments collected from real users. By adding these virtual segments to training data sets, our system achieves better system performance. Furthermore, because in practice there are variations in users' face orientations, locations, and environments (e.g., different rooms, furniture placements, etc.), we define these types of variations as domains and develop a domain adaptation framework to learn the domain-independent feature representations to improve the robustness of our system. We also propose a method to mitigate environmental impacts and eliminate static components in the mmWave signals. This lays the groundwork for adapting the *mmFit* to new environments.

The contributions of our work are summarized as follows:

- We design and implement a fitness monitoring system using a single COTS mmWave device. The proposed system integrates workout recognition, user identification, multi-user monitoring, and training effort reduction modules.
- To reduce training efforts, we develop a domain adaptation framework to reduce the amount of training data needed from different domains via mitigating impacts caused by domain characteristics embedded in mmWave signals. We also develop a GAN-assisted method to achieve better workout recognition and user identification when only limited training data from the same domain is available.
- To achieve personalized workout recognition, we propose a unique spatial-temporal heatmap feature that integrates multiple workout features, including the range of movement, velocity, and time duration. For multi-person workout monitoring, we develop a clustering-based method to derive a spatial-temporal heatmap for each user.
- We implement a prototype of *mmFit* and evaluate its performance involving 14 types of full-body workouts and over 7000 workout segments from various real-world scenarios. Our system can achieve average accuracies of 97% in workout recognition and 91% in user identification with small training data (e.g. 10 repetitions per workout type). When we only use one repetition per workout type for training, the accuracies drop moderately to 85% in workout recognition and 81% in user identification.

II. Related Work

In general, activity recognition and fitness monitoring systems can be classified into three categories: camera-based [2, 18], sensor-based [4, 5] and radio-frequency (RF) signal-based [6, 11]. In this part, we will review existing works and compare them with our proposed mmWave-based fitness monitoring system.

A couple of camera-based systems have been proposed to recognize human activities and perform fitness monitoring [3, 2, 18]. These works use cameras to capture images or videos and apply image-processing algorithms to extract motions or user identities. However, camera-based methods may raise privacy concerns. To address this weakness, sensor-based systems have been developed [19, 5, 4]. These works explore various non-intrusive sensors like gyrometer [19], ECG [5] or FSR sensor [4] to collect different types of signals for further analysis. However, sensor-based approaches require users to wear sensors or other devices, which is inconvenient for senior people or during complex activities.

To overcome the above limitations, researchers recently propose to exploit RF-based methods (e.g., WiFi and mmWave). WiFi-based approaches [6, 7] use off-the-shelf WiFi devices to infer activities and users' identifies. However, being easily influenced by surrounding environments remains the main limitation. Compared with WiFi signal, mmWave has been proven to be robust for activity recognition due to the antenna directionality and stability. Some researchers propose to build human skeleton systems to capture human postures [8, 9, 10, 20, 21]. However, these systems require special cameras (e.g., Microsoft Kinect) to provide accurate joint locations. Extra devices requirements and privacy concerns might limit the widespread deployments of these approaches for in-home fitness monitoring. Besides, many activity recognition works have shown satisfying performance for single-user scenarios. [11, 12, 13, 16, 22, 23, 24]. However, multiple family members may perform workouts simultaneously in a shared space. Thus, a system that works in concurrent scenarios is desired for fitness monitoring. To address this weakness, researchers propose approaches [17, 14, 25, 15, 26] that could track multiple people simultaneously. However, these approaches usually require users to repeat dozens of times (e.g., 30 times) of the same gesture or activity in the training stage, which is time-consuming and labor-intensive to build a module including multiple types of workouts. Furthermore, training and testing data might be different in terms of people's orientations, locations, and environments, thus requiring significant training efforts. In addition, it is still desired to have a system that could identify identities along with activities to provide personalized fitness monitoring. Thus, existing mmWave-based approaches are not suitable for in-home fitness monitoring.

Compared with existing work, our system enables low-effort personalized fitness monitoring using a single COTS mmWave device. We integrate multiple modules including workout recognition, user identification, multi-user monitoring, and training effort reduction into a single system. We develop a domain adaptation framework to reduce the efforts of training data collection from different domains and also develop a GAN-assisted method to achieve better performance. Furthermore,

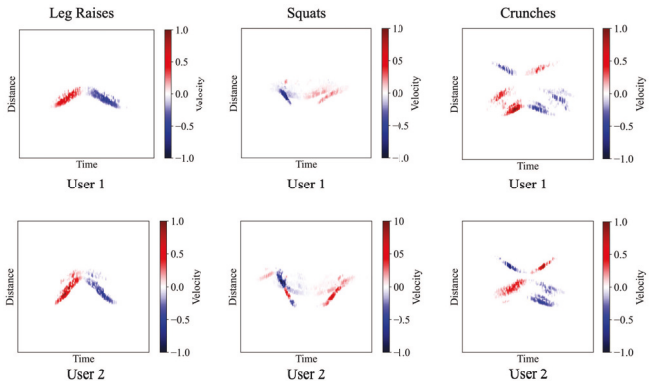


Fig. 1: Spatial-temporal heatmaps of three different workouts of two users.

we propose unique features to enable personalized workout recognition and develop a clustering-based method for multi-user monitoring.

III. Preliminaries

A. mmWave Radar Fundamental

MmWave radar system transmits frequency-modulated continuous-wave (FMCW) signals to objects with a wavelength within millimeter range [27]. The radar transmits signals called 'chirp' and receives the reflected signals from the objects. A chirp is a sinusoid whose frequency increases linearly with time. The chirp bandwidth is B and duration is Tc. The slope S of chirp defines the rate at which the chirp ramps up. The receiving signal and transmitting signal are 'mixed' and the resulting signal is called an intermediate frequency (IF) signal [28].

To facilitate fitness monitoring via mmWave, we first need to understand how a target affects the IF signal. A target object in front of the radar produces an IF signal with a constant frequency tone of $2dS/c$. The distance between the target and radar can be calculated as $\frac{f_{IF} \cdot c}{2 \cdot S}$, where c is the speed of light. Besides, the target with a moving speed v should have different phases across two consecutive chirps. The phase difference ω measured across two consecutive chirps can be used to estimate the velocity of the object through $\frac{\lambda \cdot \omega}{4\pi \cdot T_c}$, where λ is the wavelength. Furthermore, different distances from the object to different antennas on the radar result in phase differences. The measured phase difference ω across different antennas can be used to estimate the AoA of the object using $\sin^{-1}\left(\frac{\lambda \cdot \omega}{2\pi l}\right)$, where l is the distance between neighboring antennas.

B. Feasibility Study

To demonstrate the feasibility of fitness monitoring via mmWave, we conduct experiments with two participants performing different types of workouts (e.g., leg raises, squats, and crunches) in front of a mmWave device (i.e., AWR1642), respectively. Although we can measure the distance, velocity from mmWave signals separately, integrating multiple workout features to represent human dynamics is important since it exhibits the differences among various workouts more clearly. In this paper, we develop spatial-temporal heatmaps to enable personalized fitness monitoring. As shown in Figure 1, in a spatial-temporal heatmap, the horizontal axis represents the time duration of a workout repetition while the vertical axis represents the range of movement. The velocity is represented by color. We observe that spatial-temporal heatmaps of different workouts present different patterns, which demonstrate the feasibility of workout recognition. Moreover, comparing the first line and second line of Figure 1, we also observe that when different people perform the same workout, spatial-temporal heatmaps would be different due to people's various heights, strengths, and figures. This observation confirms the feasibility of using the proposed heatmaps for personalized workout fitness monitoring. The detail of building spatial-temporal heatmaps will be discussed in Section IV-B.

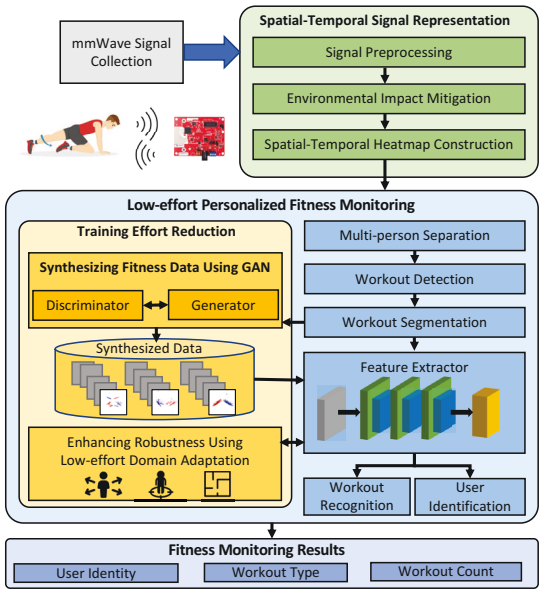


Fig. 2: System overview of mmFit.

IV. Low-effort Personalized Fitness Monitoring

A. System Overview

The main goal of *mmFit* is to perform personalized workout monitoring by examining the dynamics of mmWave signals. As shown in figure 2, our system takes as input the mmWave signals reflected from human body. It first performs signal processing to derive the velocity, distance, and AoA information of the users, respectively. Then, the system mitigates the impact from environments by subtracting signals reflected off static objects (i.e., tables and walls). To aggregate the spatial, temporal, and velocity features of human activities, we construct spatial-temporal heatmaps. Such signal representation integrates different activity characteristics (i.e., velocity, range of movements, and time duration) that facilitate both workout recognition and user identification. For multi-user monitoring, we utilize a clustering-based method to construct a spatial-temporal heatmap for each user based on the reflected signals. To detect workout activities from other daily activities, we propose a workout detection method based on the repetitive velocity pattern of workouts in time domain.

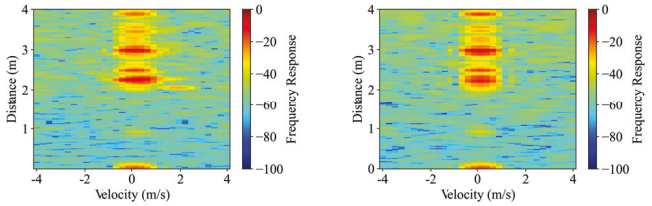


Fig. 3: (a) Doppler-range heatmap when one user is performing workout in the detection area; (b) Doppler-range heatmap when no user is in the detection area.

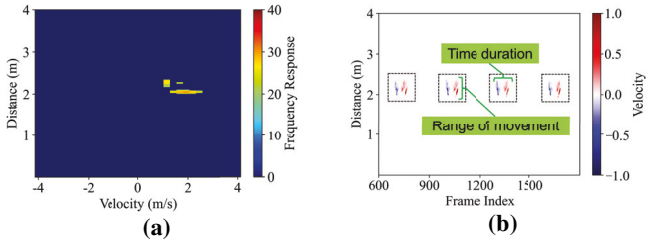


Fig. 4: (a) Doppler-range heatmap after environmental impact mitigation; (b) Spatial-temporal heatmap of four repetitions.

We develop deep learning models to perform workout recognition and user identification, respectively. An extractor is used to learn feature representations, which amplifies the distinctive workout or user characteristics. A classifier is then utilized to identify workout type or user identity. Furthermore, a domain adaptation framework is developed to handle the differences between training and testing datasets and enhance the robustness of our system. In addition, we develop a generative adversarial network (GAN) to enable better system performance when limited training data is available. We employ GAN to synthesize a large number of virtual workout segments based on a small number of segments collected from real users and improve the performance of our system by adding these virtual segments to the training dataset.

B. Spatial-Temporal Signal Representation

Signal Preprocessing. To capture fine-grained workout information, we first perform range-FFT and doppler-FFT signal processing on the raw data to derive distance and velocity measurements, respectively. Specifically, an FFT (i.e., range-FFT) is performed on the received data to convert the time domain signal into the frequency domain signals which indicate different objects with various peaks. In order to detect the velocities of objects in the sensing area, we further apply another FFT (i.e., doppler-FFT) on the range-FFT signals. After that, we could derive a doppler-range heatmap, which shows the strength of the frequency response (indicated by color) and the velocity (doppler index) of the object at a specific distance (range index) of one frame. As shown in Figure 3a, when a user is performing workouts in front of the device, the doppler-range heatmap could capture high-frequency responses. However, since static objects (e.g., furniture and walls) in the environments can also reflect mmWave signals, it is still hard to extract signals from humans.

Environmental Impact Mitigation. To mitigate the environmental interference, we propose an environmental impact mitigation method by filtering out non-moving objects in doppler-range domain. Specifically, when there is no user performing workouts in the detected area, we observe that the doppler-range heatmap of the static objects (e.g., walls and furniture) remains consistent over time as shown in Figure 3b. This motivates us to mitigate static impact by subtracting the time-invariant frequency responses. Specifically, we collect mmWave signals in a static environment for a short period (e.g., 1 min) to estimate the time-invariant frequency response from static objects. We further eliminate noise signals by removing frequency responses with low intensity from the dopper-range domain based on an empirical threshold. Through these procedures, as shown in Figure 4a, it is clear to see the frequency responses caused by human activities in the denoised doppler-range heatmap.

Spatial-Temporal Heatmap Construction. Although doppler-range heatmaps could capture velocity information in different ranges, it is insufficient to describe the process of the workout, because it does not contain temporal information, such as the time duration of each repetition and the velocity variations in time domain. To integrate multi-dimensional features, we propose to construct spatial-temporal heatmaps. Specifically, we accumulate the velocity of every distance in every doppler-range heatmap together as follows:

$$V_{q,t} = \sum_{p=1}^{D} (I_{p,q,t}) \times v_{p,t}, p \in [1, D], q \in [1, R], \qquad (1)$$

where $I_{p,q,t}$ is the intensity of a frequency response in the doppler-range heatmap, p is the doppler index, q represents the range index, and t is the frame index. $v_{p,t}$ is the velocity corresponding to a doppler index p in frame t. Then we normalize the derived $V_{q,t}$ and transfer the original instantaneous velocity-distance relationship to a more comprehensive spatial-temporal heatmap which describes the process of a workout as shown in Figure 4b.

C. Multi-User Separation

Multiple family members may perform workouts together in a shared space. In such cases, our system needs to be able to monitor multiple people's workouts concurrently using a single device. This is challenging as the mmWave signals reflected from different peoples might be mixed. We take two users' concurrent workouts as an example. When two people are performing workouts at the same distance to the mmWave device, the reflected signals from each person are hard to be differentiated in the range domain, because of the same measurements in distance. Thus, it is difficult to separate each user's workouts based on their distance in the spatial-temporal heatmap. Furthermore, when users are performing workouts at the same time, it would be even harder to separate different users' workouts because of the overlapped spatial-temporal heatmaps of two users.

We need to develop a method to separate the reflected signals from different users. We find that when people perform

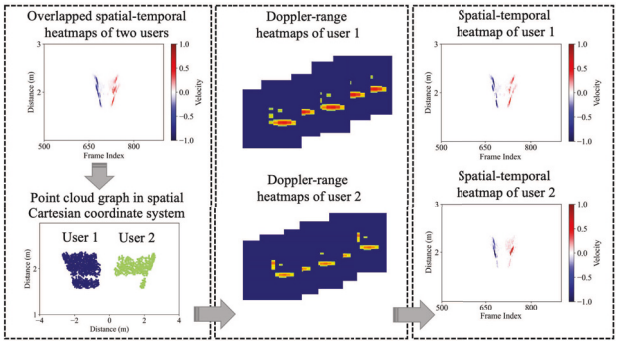


Fig. 5: The flow of multi-user separation.

workouts, each people should have their own space to allow every part of the body to move adequately. This makes the reflected signals from different people separated in the two-dimensional spatial Cartesian coordinate system. Based on this idea, in this paper, we develop a clustering-based method to achieve multiple people fitness monitoring based on different users' spatial information as shown in Figure 5. Specifically, we generate the point cloud graph in the two-dimensional spatial Cartesian coordinate system via applying an FFT (angle-FFT) on the doppler-FFT signals. In particular, given a point in the doppler-range domain with range index q, doppler index p, measured distance d, and AoA (θ), taking the position of the device as the origin, we could derive the spatial coordinate of the point as follows:

$$\begin{aligned} x_{(p,q)} &= d_{(p,q)} \times \sin\theta_{(p,q)} \\ y_{(p,q)} &= d_{(p,q)} \times \cos\theta_{(p,q)}, \end{aligned} \tag{2}$$

where x and y are x-axis and y-axis coordinate, respectively.

We accumulate the point cloud graphs of every frame in a sliding window (i.e., 10 seconds) and separate the accumulated point clouds into different clusters via an unsupervised clustering method (i.e., DBSCAN). It detects the number of users by counting the number of clusters in the detected area. When more than one cluster (the number of points in each cluster should be more than an empirical threshold P_{max}) are detected in the area, our system confirms the existence of multiple users. Next, for each cluster, we map all points in the cluster back to the doppler-range domain based on Equation 2. A spatial-temporal heatmap for each user can be derived based on the doppler-range heatmap of every frame in the sliding window using Equation 1. We note that this point cloud-based method could also be applied in more complex scenarios with more than 2 users since multiple users' reflected signals can still be detected and separated in the spatial Cartesian coordinate system. We present the evaluation results of multi-user monitoring with different group sizes in Section VI-D.

D. Workout Detection

After constructing spatial-temporal heatmaps for every user, we focus on detecting the workout from non-workout activities. There is an observation that workout activities usually have repetitive patterns in the spatial-temporal domain while non-workout activities do not. The reason is that fitness activities consist of consistent motions that usually be repeated multiple

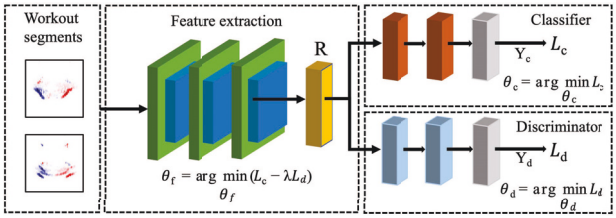


Fig. 6: Domain adaptation training framework.

times. Based on the observation, we propose to detect workouts by searching for the repetitive patterns via a sliding window. Specifically, we accumulate the velocities from all the distances in a frame to transfer the spatial-temporal heatmap to one-dimensional data. Then, an auto-correlation-based method is used to find the repeating pattern by comparing the similarity with itself at a time lag. A peak (i.e., local maximum) detection algorithm is adopted with an empirical threshold (i.e., 0.2) to derive the number of peaks N. The workout is detected within the specific sliding window (i.e., 10 seconds) when N is larger than N_{min}. We set N_{min} to be 5 in our prototype system [29].

E. Workout Clustering and Segmentation

To facilitate accurate workout recognition and user identification, we need to determine the segments of mmWave signals that only contain the repetitions of workouts. In this paper, we determine each workout repetition according to their time duration and range of movement in the spatial-temporal heatmap. Specifically, we utilize a threshold to set the points in the heatmap with low absolute velocity to zero. The threshold is selected from empirical studies. Then, we use DBSCAN to separate the points into different clusters based on the coordinates of the points in the heatmap. To segment each repetition more accurately, we design a dynamic segmentation algorithm that determines the 2D window size of each repetition according to its time duration and range of movement in the spatial-temporal heatmap. As shown in Figure 4b, for each repetition, we find the coordinates of minimum and maximum points in the frame-distance plane. Then, we derive the window size based on the differences between the coordinates of the extreme points. We scale up the size of the window by a constant (i.e., 1.2) which is empirically determined to ensure robustness.

V. Training Effort Reduction

A. Reducing Training Effort in Cross-domain Deployment

To reduce the training efforts for cross-domain deployments and improve the robustness of our system, we develop a domain adaptation framework. This framework takes the spatial-temporal segments obtained from previous processes as input. The data is first converted into a feature extractor. Based on the feature representations, a classifier is developed to predict the workout or corresponding user. To mitigate domain characteristics (e.g., user's face orientation, location, and environment) embedded in feature representations, a discriminator that can differentiate domains is developed to help optimize the feature extractor. By optimizing the feature extractor to "fool" domain discriminator but also support workout/user classifier,

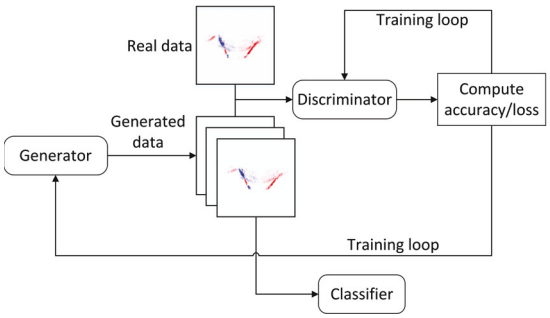


Fig. 7: Structure of the GAN for workouts generation.

the framework can facilitate the deployments across different domains and improve robustness. Figure 6 shows the structure of the proposed domain adaptation framework.

Workout Recognition and User Identification Model. The feature extractor has three convolutional layers, each with a 3×3 filter and a ReLU activation function. Each convolutional layer is followed by a max-pooling layer with a stride of 2 and a filter size of 2×2 to downsample and reduce data redundancy. After the process of 3 rounds of up-sampling and down-sampling, a 64-dimension feature map is obtained. Then, a flatten layer is integrated to reduce the feature map into a one-dimension array. Given an input data D, the feature extractor produces feature representations $R = F(D, \Theta_f)$, where F represents the feature extractor and Θ_f represents its trainable parameters. Based on the derived feature representation R, a neural network consisting of two dense layers is followed to classify the inputs into several classes (e.g., different types of workouts). We train two classifiers with same structure to differentiate workouts and user identities, respectively. Given the input representation R, the classifier predicts the label as Y_c. We optimize the classifier by minimizing the cross-entropy loss between the ground truth $\tilde{Y_c}$ and the predicted label Y_c as $L_c = L_{CE}\left(Y_c, \tilde{Y_c}\right)$, where L_{CE} represent the cross-entropy loss function.

Domain Adaptation Training. To reduce the amount of training data collected from different domains, we use domain adaptation training [30] to transfer our system to new domains. Specifically, we develop a domain discriminator to make the feature representations similar between different domains. The proposed domain discriminator consists of two dense layers. It takes the feature representations R as inputs and predicts the domain labels Y_d (e.g., original environment or a new environment). We optimize the discriminator with the cross-entropy loss between the ground truth domain $\tilde{Y_d}$ and the predicted domain Y_d as $L_d = L_{CE}\left(Y_d, \tilde{Y_d'}\right)$. To make the extracted features be domain-independent, we design an adversarial loss as $L_{adv} = L_c - \lambda L_d$, where L_c is the classification loss and L_d is the discriminator loss. We set domain loss L_d to negative since we would like to train the feature extractor to maximize the domain loss. The factor λ is selected to control the balance between the transferability and distinguishability of the extracted features.

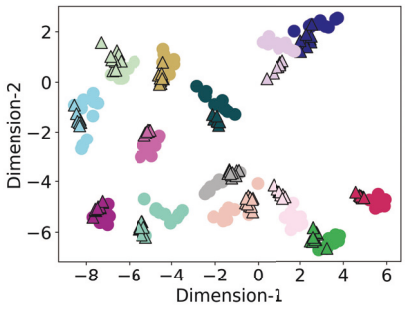


Fig. 8: Distributions of real (circle) and generated (triangle) workouts. Each color represents one workout type.

B. Reducing Training Effort Using GAN

We develop a GAN-assisted method to achieve better user identification and workout recognition when only limited training data from the same domain is available. GAN has been exploited in a wide range of scenarios like human identification [31] or smart medicare [32, 33] with promising results. One function of GAN is generating synthesized data to reduce training effort. In our work, we seek to use GAN to extend workout training data.

Workout Data Generation Using GAN. The GAN used in our work is composed of two sub-networks: a generator G and a discriminator D. As shown in Figure 7, G aims to learn the distribution of the real workout segments to synthesize virtual data. D tries to discriminate whether a workout is real or synthesized. G and D compete with each other to achieve their goals, which leads to a two-player mini-max game. The overall value function $V(D, G)$ is defined by:

$$\min_G \max_D V(D, G) = \log(D(x)) + \log(1 - D(G(n))), \quad (3)$$

where x represents a real workout, n represents a noise vector. $G(n)$ is synthesized workout generated by G. $D(\cdot)$ is discriminator's estimation that the data is a real workout. The basic training process of GAN in *mmFit* is described as follows: G tries to maximize the probability that a synthesized workout is discriminated as real by minimizing the generator loss function $L_G = -\log(D(G(n)))$. On the other hand, D tries to maximize the probability that the real and virtual workout are classified correctly by maximizing the discriminator loss $L_D = -\log(D(x)) - \log(1 - D(G(n)))$. G and D are trained in turn to optimize each other by updating parameters of their networks. The final state is a Nash equilibrium where the generated workouts are similar to the real ones, and the discriminator fails to identify whether the workouts are real or not.

GAN Implementation and Visualization. In *mmFit*, both G and D are implemented by convolutional neural networks (CNN). G has 5 transpose convolutional layers. D has 6 convolutional layers followed by a flatten layer. After the competing of the generator and the discriminator, our GAN network eventually generate a large number of high-quality synthesized workouts. Even users only perform a small number of repetitions in real-life scenarios, with the assistance of the generated workout dataset, *mmFit* can be well-trained offline

and thus guarantee a good workout recognition and user identification performance online. To give an overall evaluation of the quality of the GAN-generated workouts, we generate virtual workouts based on a small number of real workouts (i.e., one repetition per workout type). Then we utilize t-SNE [34] to visualize the similarity between real and GAN-generated workouts. Specifically, GAN-generated and real workouts are firstly converted by the feature extractor as discussed in Section V-A. Then, t-SNE is applied to the extracted high-dimensional feature representations for dimension reduction. Finally, the feature embedding of each workout is obtained and illustrated in a 2-D plane as shown in figure 8. We observe that the feature embedding of the GAN-generated repetitions overlap with the real ones from the same workout type but have no overlap with those from different types. This indicates that the GAN-generated repetitions are similar to real repetitions and can increase the insufficiency of workout training data. We further validate the effectiveness of our GAN-assisted method in section VI.

VI. Performance Evaluation

A. Experimental Setup

Devices. We build a prototype of *mmFit* using a single mmWave device, TI AWR1642 [35], which integrates a 2×4 antenna array. The device operates in a frequency range between $77GHz$ and $81GHz$. The sampling rate is fixed at 100 frames per second and each frame has 17 chirps. A TI DCA1000EVM [36] data capture card is used to collect data from the mmWave device and send data to a dell laptop for deep model inference.

Data Collection. We recruit 11 volunteers aged from 20 to 44 with various heights from $162cm$ to $185cm$ and weights from $50kg$ to $86kg$. The volunteers are asked to perform 14 typical indoor workouts as shown in Table I. The workout data are collected from four different environments (e.g., lounge, corridor, and classroom). We place the mmWave device on a table with a height of $60cm$ and the ground truth videos are recorded by a camera. For multi-user scenarios, volunteers are randomly separated into different groups with size from 2 people to 4 people to perform workouts concurrently. During the workouts, all the volunteers are required to keep a reasonable distance from others (e.g., $3m$) to avoid physical touch. During an eight-month survey, we ask each of the volunteers to conduct 20 repetitions for each type of workout (Unless mentioned otherwise, 10 segments are used for training the model, and 10 segments are used to evaluate the performance). In total, we collect over 7000 segments from the volunteers.

Evaluation Metrics. To evaluate the performance of our system, we define the following evaluation metrics: *workout recognition/user identification accuracy* is the probability of workouts/user identities that are correctly recognized by *mmFit*; *confusion matrix* visualizes the percentage of each workout been identified among all workouts (i.e., the correct workout and the other workouts).

TABLE I: 14 common in-home full-body workouts

W1	Crunches	W8	Squats
W2	Elbow plank and reach	W9	Burpees
W3	Leg raise	W10	High knees
W4	Lunges	W11	Turning kicks
W5	Mountain climber	W12	Chest squeezes
W6	Punches	W13	Side leg raise
W7	Push ups	W14	Side to side chops

B. Workout Recognition

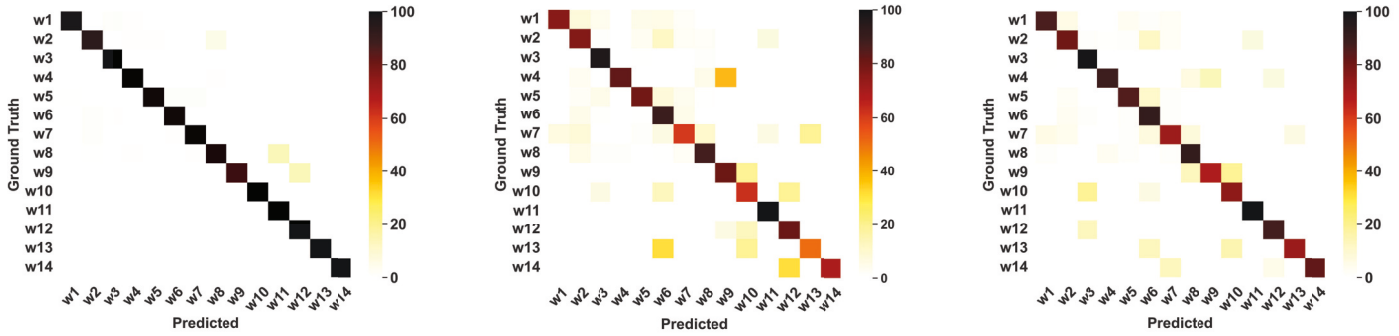
We first examine the performance of our system for workout recognition. As demonstrated in Figure 9a, when training data is sufficient (i.e., 10 repetitions per workout per user), our system can achieve an average workout recognition accuracy of 97.69%. Moreover, the recognition accuracy of all workout types is above 87.5%. However, in real-life scenarios, it is usually not practical to ask every user to perform large amounts of workout repetitions of every workout type. For example, when people could only provide a limited amount of repetition (e.g., one repetition per workout type), a workout recognition system usually could not achieve good performance. In such condition, as shown in Figure 9b, the system's average workout recognition performance is 80%, and the average recognition accuracy for some workout types (i.e., w7, w10, and w13) is even lower than 65%.

To deal with such conditions, we utilize GAN to generate a large number of virtual workout segments based on the limited amount of real workout data and utilize these segments to improve the capability of our system. As shown in Figure 9c, with the GAN-generated workout segments, the workout recognition system could achieve better performance with an average accuracy of 85%. We also notice that the recognition performance for all 14 workouts is increased by at least 3%. Some workouts (w3, w6 $\&$ w8) reach a performance of higher than 90% recognition accuracy. This result not only demonstrates that our system could achieve satisfying workout recognition performance with a small amount of real workout training data, but also proves the effectiveness of the proposed GAN-assisted method.

We further explore the minimum number of GAN-generated segments that the system needed to achieve satisfactory performance. Figure 10 shows that the average workout recognition accuracy (marked by red points) gradually increases from 74% to 79% when the number of GAN-generated segments of each workout category increases from 1 to 9. However, when more than 9 generated segments are added to the training dataset, the overall accuracy has no obvious improvement. The result demonstrates that 9 GAN-generated repetitions for each workout/user category is sufficient for our system to achieve obvious performance improvements, thus we use 9 GAN-generated repetitions as our default settings.

C. User Identification

We then examine the performance of our system for user identification. As is shown in Figure 11a, when training data is sufficient, the user recognition accuracy of all volunteers is



(a) (b) (c)

Fig. 9: (a) Workout recognition performance of our system with sufficient training data (e.g., 10 repetitions per workout); (b) Workout recognition performance without GAN-assisted method when only one real repetition per workout is available; (c) Workout recognition performance with GAN-assisted method.

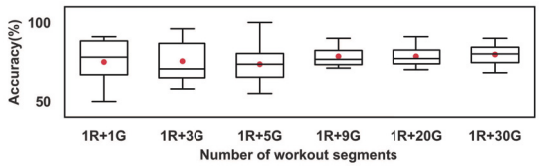
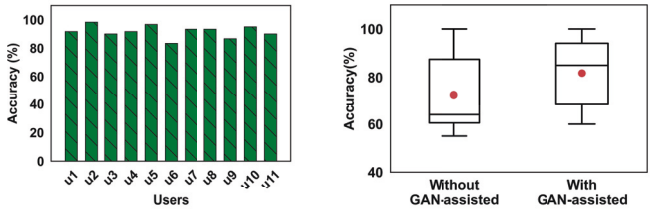


Fig. 10: Impact of the number of workout segments in training set (e.g., $1R + 9G$ represents 1 real workout repetition with 9 synthesized repetitions).

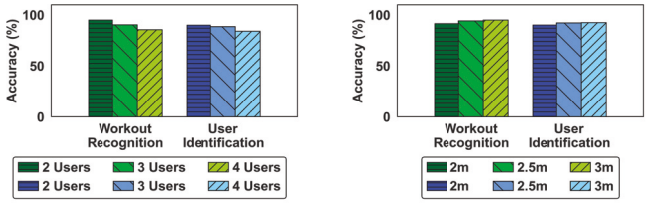


(a) (b)

Fig. 11: (a) User recognition of our system with sufficient repetition training data (e.g., 10 repetitions per workout); (b) User recognition without and with our GAN-assisted method when only one repetition per workout is available.

higher than 83%. The average accuracy of all users is 92% and the highest accuracy can reach up to 98.33% (U2). We also notice that the performances for some users (e.g., U6) are slightly lower than others, this is because this user has similar body shape and workout patterns with other users. This result shows that our system can identify different users with high accuracy and thus supports personalized workout recognition.

Similar to workout recognition, we further explore the user identification performance of our system when only limited workout data is available. Specifically, we utilize only one repetition per workout per user to train the user classifier. In such conditions, the system only achieves an average user recognition performance accuracy of 72% (marked by the red point) as shown in Figure 11b. However, with the proposed GAN-assisted method, the user identification performance of our system is boosted to an average accuracy of 81%. We also notice that the highest, median, and quartile values of user identification accuracy are all improved over 5%. These results demonstrate that our GAN-assist method has the potential to



(a) (b)

Fig. 12: (a) Workout recognition and user identification performance in multi-person scenarios with sufficient training data. (b) Impact of distance among users on workout recognition and user identification performance.

be utilized in other RF-based recognition systems to achieve better performance.

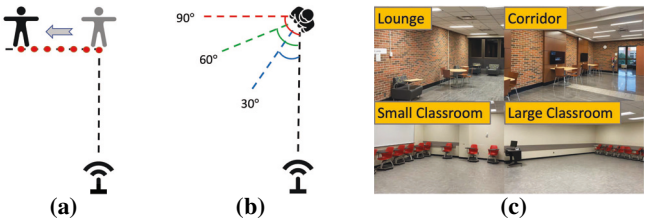
D. Multi-person Monitoring

Multiple family members might perform workouts concurrently at home. Our system enables multi-user monitoring by using a clustering-based method as discussed in Section IV-C. To evaluate the performance of our system for multi-user fitness monitoring, we ask multiple people with different group sizes (i.e., 2, 3, and 4) to perform workouts and present the result in Figure 12a. Specifically, in two-user cases, when sufficient training data is available, our system could achieve a workout recognition accuracy of 95% and a user identification accuracy of 90%. When group size increase to three people, our system could achieve a workout recognition accuracy of 90.3% and user identification accuracy of 88.57%. Even when there are four people performing workouts concurrently, our system can still achieve workout recognition accuracy of 85.55% and user identification accuracy of 84%. This result demonstrates that our fitness monitoring system can be applied in family scenarios where generally there will be no more than 4 persons performing workouts together.

Moreover, when family members could only provide a limited amount of repetitions, with the help of our GAN-assisted method, we still observe obvious performance improvements of our system in different scenarios. As shown in Table II, the workout recognition and user identification accuracy in three scenarios have an average improvement of 5.83% and

TABLE II: The performance of workout recognition and user identification in multi-person scenarios when only one repetition per workout training data is available.

	Two Users	Three Users	Four Users	Average Improvement
Workout Recognition without GAN-assisted	72.73%	69.62%	60.55%	**5.83%**
Workout Recognition with GAN-assisted	77.00%	73.15%	70.25%	
User Identification without GAN-assisted	69.56%	69.18%	58.25%	**6.12%**
User Identification with GAN-assisted	75.93%	71.66%	67.75%	



(a) (b) (c)

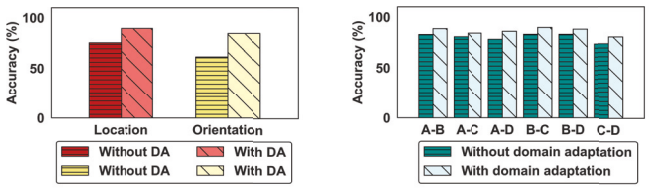
Fig. 13: (a) Different relative locations between the user and the device; (b) Different face orientations relative to the device; (c) Different environments between the training and testing datasets.

6.12%, respectively. This result proves that our GAN-assisted method facilitates the practicability of applying our system for family workout monitoring with low training efforts. We further evaluate the impact of distance among users in multi-person scenarios. We ask two volunteers to perform workouts concurrently with a distance of $2m$, $2.5m$, and $3m$ between each other, respectively. As shown in Figure 12b, our system achieves higher than 90% accuracy in both workout recognition and user identification for all 3 scenarios. Note that the distance among users is selected based on the workout type and the users' heights to avoid touching each other.

E. Performance of Domain-independent Training cross Different Locations/Orientations

People might have different locations relative to the devices between the training and testing datasets. We first evaluate the transferability of our system to different locations. Specifically, we set the point in front of the device with a distance of $2.5m$ as the original point. We collect workout data from 6 new locations (i.e., each has a distance ranging from $15cm$ to $90cm$ to the original point) with an interval of $15cm$ as shown in Figure 13a. We use workouts collected at original point as the source domain and all the workouts collected at other locations as the target domain. In addition, to evaluate the transferability of our system to users' different face orientations, we also set the direction that faces the device as the original orientation. The volunteers are asked to perform workouts at the same location with 7 different face orientations (i.e., $0°$ to $90°$ with an interval of $15°$) as shown in Figure 13b. Workouts collected at $0°$ are used as the source domain while workouts collected at other orientations are chosen as the target domain.

We evaluate the performance of *mmFit* via training the deep learning model with workouts collected from the source domain



(a) (b)

Fig. 14: (a) Workout recognition without and with domain adaptation (DA) across different relative locations and face orientations; (b) Workout recognition without and with domain adaptation across different environments (e.g., A-B represents training data from Environment A and testing data from Environment B).

and adapt the model to the target domain using the domain adaptation framework described in Section V-A. For comparison, we also examined the deep learning model without domain adaptation. As shown in Figure 14a, we observe that without domain adaptation, the performance of workout recognition is low (i.e., 75.25% across different locations and 61.37% across various orientations). Our domain adaptation framework boosts the system's workout recognition performance on new locations and orientations to 89.5% and 84.45%, respectively. The result demonstrates that the proposed domain adaptation method can alleviate the impact of different locations/orientations between training and testing data on our system.

F. Performance of Domain-independent Training cross Different Environments

To evaluate the transferability of our system to different environments, we collect workout data from 4 different rooms with various furniture layouts: a lounge with a size of 7×4 m^2, a corridor with a size of 10×4 m^2, and two classrooms with the size of 6×8 m^2 and 10×10 m^2, respectively, as shown in Figure13c. In each environment, the data are collected with the same location and face orientation relative to the device. Workouts collected in one environment is chosen as the source domain while workouts collected in another environment is selected as the target domain. We test all 6 source-target evaluation pairs and present the result in Figure 14b. We observe that with the help of domain adaptation framework, our system achieves 86% workout recognition accuracy with an average 4% improvement compared with not using domain adaptation. We also note that our system still achieves good transferability across different environments even without using domain adaptation training which validates the effectiveness of our environmental impact mitigation methods as discussed in Section IV-B. The above results show that our system can be easily deployed across different room environments, which is essential to the practical fitness system deployment in everyday life.

VII. Conclusion

In this paper, we propose a novel fitness monitoring system using a single COTS mmWave device. This system integrates workout recognition, user identification, multi-user monitoring, and training effort reduction modules and makes them work together in a single system. To reduce training efforts, we a

develop domain adaptation framework to reduce the amount of training data collected from different domains via mitigating impacts caused by domain characteristics embedded in mmWave signals. We also develop a GAN-assisted method to achieve better user identification and workout recognition when only limited training data from the same domain is available. To achieve personalized workout recognition, we propose a unique spatial-temporal heatmap feature to integrate multiple workout features. We also develop a clustering-based method for concurrent workout monitoring. Extensive experiments with 14 typical workouts involving 11 participants demonstrate that our system can achieve 85% and 81% accuracy for workout recognition and user identification using only one sample of each workout/user.

VIII. Acknowledgment

This work was partially supported by the National Science Foundation Grants CCF-2028873, CCF-1909963, CNS-2120350, CCF-2000480, CCF-2028858, CNS-2120276, CNS-2145389, CNS-2120371, CCF-2028894, CNS-1717356, CNS-2120396, CCF-2028876, CCF-1909963, ECCS-2033433.

References

[1] "Lack of physical activity," https://www.cdc.gov/chronicdisease/resources/publications/factsheets/physical-activity.htm, 2019.

[2] O. Celiktutan, C. B. Akgul, C. Wolf, and B. Sankur, "Graph-based analysis of physical exercise actions," in *Proceedings of the 1st ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare*, ser. MIIRH '13. New York, NY, USA: Association for Computing Machinery, 2013.

[3] E. Ghorbel, R. Boutteau, J. Boonaert, X. Savatier, and S. Lecoeuche, "Kinematic spline curves: A temporal invariant descriptor for fast action recognition," *Image and Vision Computing*, 2018.

[4] E. A. H. Akpa, M. Fujiwara, Y. Arakawa, H. Suwa, and K. Yasumoto, "Gift: Glove for indoor fitness tracking system," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2018.

[5] Y. Meng, S.-H. Yi, and H.-C. Kim, "Health and wellness monitoring using intelligent sensing technique," *JIPS*, 2019.

[6] X. Guo, J. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah, "Device-free personalized fitness assistant using wifi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018.

[7] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained wifi signatures," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '14. New York, NY, USA: Association for Computing Machinery, 2014.

[8] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-pose: Real-time human skeletal posture estimation using mmwave radars and cnns," *IEEE Sensors Journal*, 2020.

[9] H. Xue, Y. Ju, C. Miao, Y. Wang, S. Wang, A. Zhang, and L. Su, "mmmesh: Towards 3d real-time dynamic human mesh construction using millimeter-wave," in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021.

[10] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "Rf-based 3d skeletons," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, 2018.

[11] G. Tiwari and S. Gupta, "An mmwave radar based real-time contactless fitness tracker using deep cnns," *IEEE Sensors Journal*, 2021.

[12] P. S. Santhalingam, A. A. Hosain, D. Zhang, P. Pathak, H. Rangwala, and R. Kushalnagar, "mmasl: Environment-independent asl gesture recognition using 60 ghz millimeter-wave signals," *Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020.

[13] Y. Wang, H. Liu, K. Cui, A. Zhou, W. Li, and H. Ma, "m-activity: Accurate and real-time human activity recognition via millimeter wave radar," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.

[14] M. A. U. Alam, M. M. Rahman, and J. Q. Widberg, "Palmar: Towards adaptive multi-inhabitant activity recognition in point-cloud technology," *arXiv preprint arXiv:2106.11902*, 2021.

[15] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-wave micro-doppler signatures," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[16] Y. Ren, J. Lu, A. Beletchi, Y. Huang, I. Karmanov, D. Fontijne, C. Patel, and H. Xu, "Hand gesture recognition using 802.11 ad mmwave sensor in the mobile device," in *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, 2021.

[17] H. Liu, Y. Wang, A. Zhou, H. He, W. Wang, K. Wang, P. Pan, Y. Lu, L. Liu, and H. Ma, "Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020.

[18] C.-J. Su, C.-Y. Chiang, and J.-Y. Huang, "Kinect-enabled home-based rehabilitation system using dynamic time warping and fuzzy logic," *Applied Soft Computing*, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494614001859

[19] M. M. Hassana, Z. Uddin, A. Mohamed, and A. Almogrena, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, 2018.

[20] M. Zhao, Y. Liu, A. Raghu, T. Li, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human mesh recovery using radio signals," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019.

[21] M. Zhao, T. Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[22] J. Wang, L. Zhang, C. Wang, X. Ma, Q. Gao, and B. Lin, "Device-free human gesture recognition with generative adversarial networks," *IEEE Internet of Things Journal*, 2020.

[23] F. Jin, A. Sengupta, and S. Cao, "mmfall: Fall detection using 4-d mmwave radar and a hybrid variational rnn autoencoder," *IEEE Transactions on Automation Science and Engineering*, 2020.

[24] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: Association for Computing Machinery, 2018.

[25] C. Wu, F. Zhang, B. Wang, and K. R. Liu, "mmtrack: Passive multi-person localization using commodity millimeter wave radio," in *IEEE INFOCOM 2020*. IEEE, 2020.

[26] F. Jin, R. Zhang, A. Sengupta, S. Cao, S. Hariri, N. K. Agarwal, and S. K. Agarwal, "Multiple patients behavior detection in real-time using mmwave radar and deep cnns," in *2019 IEEE Radar Conference (RadarConf)*. IEEE, 2019.

[27] "Frequency-modulated continuous-wave radar (fmcw radar)." [Online]. Available: https://www.radartutorial.eu

[28] F. Langford-Smith *et al.*, "Radiotron designer's handbook," 1941.

[29] "Weight training: How many reps (and sets) to do." [Online]. Available: https://www.dummies.com/health/exercise/weights/weight-training-how-many-reps-and-sets-to-do/

[30] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The journal of machine learning research*, 2016.

[31] X. Qian, Y. Fu, T. Xiang, W. Wang, J. Qiu, Y. Wu, Y.-G. Jiang, and X. Xue, "Pose-normalized image generation for person re-identification," in *Proceedings of the European conference on computer vision*, 2018.

[32] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using gan for improved liver lesion classification," 04 2018.

[33] M. Moradi, A. Madani, A. Karargyris, and T. Syeda-Mahmood, "Chest x-ray generation and data augmentation for cardiovascular abnormality classification," 03 2018.

[34] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, 2008. [Online]. Available: http://jmlr.org/papers/v9/vandermaaten08a.html

[35] A. A. Pirkani, S. Pooni, and M. Cherniakov, "Implementation of mimo beamforming on an ots fmcw automotive radar," 06 2019.

[36] L.-Z. Zhenyu, W. Lu, J. Wu, S. Yang, and G. Li, "A pelt-kcn algorithm for fmcw radar interference suppression based on signal reconstruction," *IEEE Access*, 02 2020.