Energy & Buildings 249 (2021) 111228

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

Enabling non-intrusive occupant activity modeling using WiFi signals and a generative adversarial network

Qizhen Zhou [a], Jianchun Xing [a,*], Qiliang Yang [a], Xu Wang [b], Wenjie Chen [a], Yixin Mo [a], Bowei Feng [a]

[a] College of Defense Engineering, Army Engineering University of PLA, No. 1, Haifu Xiang, Nanjing 210007, China
[b] School of Software, Tsinghua University, No. 1, Qinghuayuan, Beijing 100084, China

ARTICLE INFO

Article history:
Received 19 December 2020
Revised 16 June 2021
Accepted 17 June 2021
Available online 26 June 2021

Keywords:
Occupant activity modeling
Wi-Fi
Intelligent building
Generative adversarial network
Internet of Things

ABSTRACT

Occupant activity (OA) is a crucial prerequisite for providing energy-efficient and occupant-centric services in intelligent buildings. To understand OAs well, conventional technologies require either wearable equipment of smart devices or the deployment of specialized cameras, raising relevant issues of body and privacy-intrusion. Recent WiFi-based methods circumvent the above issues and characterize the OA patterns in a non-intrusive manner. However, they demand the cooperation of occupants for a sufficient amount of collected samples, otherwise, the performance might degrade significantly. In this paper, we propose a non-intrusive approach that models distinct OA patterns for comprehensive understanding. Based on the idea of generative adversarial network (GAN), our technical novelties are three-fold. First, we modify the working principle of vanilla GAN with external constraints to avoid brute-force generation. Second, we integrate a self-attention mechanism to establish contextual relations from both local and global perspectives. Third, to recover the informative details, we construct a powerful generator with deep residual convolutional operations. We conduct extensive experiments on a real dataset and visualize the generation results for intuitive evaluation. Numerical results compared with several state-of-the-art baselines further illustrate the superior performance of our proposed GAN approach.

1. Introduction

Worldwide, approximately four-fifths of people's entire lives are spent indoors, causing more than two-fifths of the energy resources to be consumed in the building sector [1,2]. Therefore, it is essential to promote intelligent buildings for both human-centered adaption and energy-efficient operation [3]. A building management system (BMS) is a key enabler for intelligent buildings, which helps in controlling, monitoring and optimizing building services [4]. To ensure that the BMS operates well, the fine-grained perception and comprehensive understanding towards occupant activity (OA) are both utmost [5]. This is because by modeling occupants' interactions with buildings and involving OA patterns with the BMS, an enormous variety of high-quality services regarding users' comfort and energy conservation can be realized [6], such as automatic lighting adjustment [7], thermal comfort control [8], and residents' health management [9].

To model OA patterns, mainstream works leverage either specialized wearable devices (*e.g.*, accelerometers [10] and smart phones [5]) or depth cameras [11]. Such approaches achieve high performance under certain circumstances, however, most of them are inevitably intrusive. For example, wearables can function only when the devices are properly carried by or are attached to human bodies; cameras can record OA patterns for real-time analysis, while raising the risk of privacy invasion. In this paper, we aim to eliminate the intrusions of devices and provide a transparent sensing paradigm for OA. With advancements in the Internet-of-Things (IoT), it seems appealing to use invisible WiFi signals for non-intrusive OA recognition (OAR) [12]. On the one hand, pervasive WiFi signals offered by existing WiFi infrastructure can be reused to depict OA spatial–temporal properties caused by body reflection, refraction and shadowing. On the other hand, the proliferation of WiFi-enabled IoT facilities holds promise for dense WiFi links and wide monitoring coverage [13]. Significant progress has been made in exploiting fine-grained WiFi data, more specifically, channel state information (CSI), for a wide range of regular context-aware services [7,14–16]. These methods require profound knowledge to process OA-related signal variations. However, OA patterns are changeable and highly complex due to the dynamic wireless environment [17].

* Corresponding author.
E-mail address: xjc@893.com.cn (J. Xing).

https://doi.org/10.1016/j.enbuild.2021.111228
0378-7788/© 2021 Elsevier B.V. All rights reserved.

Recent studies based on deep learning are flourishing, and shed light on the potential for robust OA modeling with less manual effort. Notable efforts have been devoted to designing proper structures for IoT applications [18], adapting deep neural networks (*e.g.*, convolutional neural network (CNN) and recurrent neural network (RNN)) for specific WiFi sensing tasks [19,20], and combining diverse learning strategies for general usage [21]. These methods learn from a large amount of labeled data to guarantee the performance, however, it is always impractical to include all kinds of OA patterns, such as variations induced by different body shapes, performing habits, experimental layouts in a one-off manner [22]. In addition, elaborate data collection tasks require the active cooperation of users, which can greatly impact users' experience and violate the principle of non-intrusive detection. Common approaches augment self-collected datasets via geometric transformations, which do little to help improve the sample diversity. Some researchers have attempted to learn transferable knowledge from public datasets [23–25], and guide the prediction of OA in the target domain [26]. However, it is non-trivial to align diverse feature subspaces under noisy circumstances. Emerging generative adversarial networks (GANs) enable the vision of non-invasive modeling, which can emulate the implicit distribution of realistic OA data through a zero-sum game [27,28]. However, unsupervised GANs cannot be controlled to produce multi-class samples with specified attributes. In addition, vanilla extensions of GAN require complex modifications to retrieve detailed OA content and optimize the training progress.

To overcome the above limitations, we present a conditional self-attention GAN framework for general WiFi-based OA modeling, called OA-GAN. On the basis of a convolutional generator-discriminator network, our main technical innovations are threefold. First, we condition the unsupervised model on external constraints for directional outputs. Driven by auxiliary class labels, the generator creates candidate samples from a latent space resembling real OA variations with the same label, and the discriminator attempts to distinguish the synthetic instances from batches of labeled data for authenticity. Second, we introduce a self-attention mechanism for both local and global content generation [29]. Since OAs are spatial–temporal events and WiFi signals embody these clues from local and global perspectives, we stack the convolutional layers for multi-scale local features, and jointly capture the short/long-range dependencies regardless of the space–time distances between elements. Third, we improve the generation efficiency by using residual convolutional module [30]. As the generator may produce a high training loss and may be easily overwhelmed by the discriminator due to vanishing gradients, we construct a stronger generator with the gradient super-highways, *i.e.*, shortcut connections, for unhindered information flow and powerful representation. We further conduct extensive experiments on a real OA dataset for visualization and authenticity evaluation. A comparison with some state-of-the-art (SOTA) approaches demonstrate the superiority of our proposed GAN framework.

Our contributions are summarized as follows:

- We enable non-intrusive OA modeling for occupant-driven building services using in-air WiFi signals and powerful GAN framework.
- We retrofit vanilla GAN for WiFi-based OA modeling by taking advantage of auxiliary constraints, local–global dependencies and deep residual convolutional operations.
- We verify the efficiency of the proposed GAN framework by conducting extensive experiments on the real dataset and comparing the performance with SOTA baselines.

The remainder of the paper is organized as follows. Section 2 illustrates the importance of OA for intelligent buildings and reviews the SOTA literature in WiFi-based OAR that provides a non-intrusive detection method. Section 3 describes the preliminaries, followed by the introduction of technical details in Section 4, the experimental evaluation in Section 5 as well as the discussion and limitations in Section 6. In the last section, Section 7, we summarize the content of this paper.

2. Literature review

In this section, we present a brief review with regard to the importance of OA for intelligent buildings and the recent progress on non-intrusive WiFi-based OA.

2.1. Importance of OA for intelligent buildings

In the human-centered era, the primary goal of the intelligent building is to understand the demands of occupants and provide the optimal solutions for users' satisfaction [5]. To achieve this goal, intelligent buildings should take the occupant-building interactions (OBIs) into account, allowing for automatic adaptions on thermal, visual, acoustic, and air quality controls [3]. In other words, the knowledge of OA determines the large variability affecting occupant experience during the building operations [6]. For example, Alex may prefer an enclosed workplace when he is absorbed in debugging, while changing his mind for an open, light and a well-ventilated area when he shares his code with colleagues. OA-driven applications should respond to the user's intention in a timely manner, and adaptively refine the settings. However, comprehensive operations are required to provide a generic solution. For example, diverse OA patterns may be induced when Alex performs the same activity in various working environments, which would confuse the BMS with improper feedback. We envision an intelligent building that can learn invariant OA properties from existing data sources and project them into the new scenarios.

In addition, OA is an essential prerequisite for the design and control strategies of energy-intelligent buildings[4,10]. It is reported that building accounts for 41% of the total energy use in the U.S., approximately half of which is consumed by the use of demand-driven appliances [31]. If an intelligent building can fulfill the consumption demand with the efficient use of building units, an enormous proportion of the building energy cost will be saved [32]. For example, occupant-adaptive management of lighting systems can contribute to a 35–75% reduction in energy usage in buildings [33]. However, most OA-enabled intelligent buildings can detect only simple and coarse-grained occupant motion states, such as occupant density and counting passers-by. With the evolution of IoT and information technology, we seek fine-grained and complex OA modeling to support flexible energy-saving services.

Understanding OA is also of crucial importance for security management, such as detecting intrusion, alerting emergency services and monitoring the health of elderly people [34]. However, common methods (*e.g.*, wearables and cameras) may fail to provide 24–7 monitoring in certain places, for example, slip-and-fall accidents occurring in the bathroom. In addition, it is impractical to collect sufficient samples for various types of "abnormal events". We develop a user-friendly approach that enlarges the OA dataset with less disturbance.

2.2. Non-intrusive WiFi-based OAR

Numerous studies have demonstrated that WiFi can support non-intrusive, fine-grained and all-around OA sensing. Early attempts at WiFi-based OAR were devoted to feature-based approaches that rely on the guidance of experts [23]. For example, Zhou et al., [15] designed specific de-noising algorithms for signal recovery in office settings. Wu et al. revealed the essence of gait cycles through delicate signal transformations [14]. To alleviate human interference in feature design, pioneers explored the potential of automatic feature learning and applied SOTA deep learning networks. For example, Yao et al., [18] first proposed a unified deep framework for automatic classification of mobile data. Zou et al., [13] presented a CNN-RNN model that can process OA patterns in a one-off way without any prior constraints. Ma et al., [24] retrofitted a CNN and improved the granularity of gesture signals, which promised natural OBIs manner for users. Sheng et al., [26] learned a deep transferrable network for cross-domain OA identification. However, these powerful models may encounter the performance degradation when there are only a few labeled data points for training.

With the advent of GANs [27], illuminating works have tried to generate OA dynamics through competitive learning. Chen et al., [17] extended the GAN prototype for occupant-driven building services. Li et al., [35] augmented the diversity of location fingerprints with a deep convolutional GAN (DCGAN). Wang et al., [28] constructed visual radio images using an iterative image-to-image CNN. Xiao et al., [22] transferred the styles of WiFi-based variations for left-out users through a dual-learning-based GAN. However, GANs and their updated version (DCGANs [36]) commonly have the limitations of unstable training and nonsensical generation [37]. In this paper, we pursue some breakthroughs in this aspect, *i.e.*, stabilizing the training progress and producing semantically meaningful outputs for occupant-driven services.

3. Preliminaries

In this section, we will briefly review the background knowledge of WiFi CSI and the fundamental theories of DCGANs.

3.1. WiFi CSI

Owing to the release of CSI Tool [40], CSI is now easily extractable from any modern WiFi-enabled facilities equipped with commercial WiFi network interface cards (NICs). To leverage the CSI for OA modeling, we should understand beforehand what CSI is and how CSI reflects entity motions in measurements. CSI is collected from each received WiFi packet, which is the sampled version of channel response in time–frequency space. In transmission schemes using Orthogonal Frequency Division Multiplex (OFDM) technology, CSI is recorded in the format of a complex vector, depicting both the amplitude and phase variations of the subcarriers between a pair of WiFi devices [25]. In the IEEE 802.11n standard, there are 56/114 subcarriers in a 20/40 MHz channel, while commercial WiFi card (*i.e*, Intel 5300 NIC) can report only 30 subcarriers that evenly spread in the channel. Nevertheless, such CSI measurements continuously collected from multiple transmit-receive antenna pairs are sufficient to reflect the properties of OA-related signal fluctuations in the spatio-temporal domain, and thus widely used in wireless sensing.

Specifically, let $H^t_{s,m}$ indicates the channel estimation for subcarrier s in m-th antenna pair at time t, and let $y^t_{s,m}$ and $x^t_{s,m}$ denote the transient signals transmitted from $N_{\rm tx}$ antennas and received by $N_{\rm rx}$ antennas, respectively. Then we can deduce the CSI value $H^t_{s,m}$ from following equation: $y^t_{s,m} = H^t_{s,m} \times x^t_{s,m} + Noise$, where $s \in$ [1,30] and $m \in [1,\ N_{\rm tx} \times N_{\rm rx}]$. The CSI value $H^t_{s,m}$ can also be expressed by the amplitude $||H^t_{s,m}||$ and phase information $\angle H^t_{s,m}$ $=||H^t_{s,m}||\exp(\angle H^t_{s,m})$. Due to hardware imperfections in commercial WiFi NICs, the phase information of CSI are vulnerable and easily affected by packet detection delay, sampling frequency offsets (SFO) and central frequency offsets (CFO) between transmit-receive antenna pairs. Therefore, we abandon the use of phase information and model the OA using only CSI amplitude. Since OA are spatial–temporal events, we construct multi-dimensional radio images, which integrate the relatively stable amplitude information A of 30 subcarriers in $N_{\rm tx} \times N_{\rm rx}$ antenna pairs and associate the successive k CSI samples as:

$$A_{30\times k\times(N{\rm tx}\times N{\rm rx})} = \begin{bmatrix} \| H^1_{1,m} \| & \cdots & \| H^j_{1,m} \| & \cdots & \| H^k_{1,m} \| \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \| H^1_{s,m} \| & \cdots & \| H^j_{s,m} \| & \cdots & \| H^k_{s,m} \| \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \| H^1_{30,m} \| & \cdots & \| H^j_{30,m} \| & \cdots & \| H^k_{30,m} \| \end{bmatrix}_{30\times k\times(N{\rm tx}\times N{\rm rx})} \tag{1}$$

For the same activity, different pairs of antennas as well as subcarriers offered by the same receiving antenna can observe different CSI fluctuations due to spatial diversity, while successive CSI measurements are able to capture the contextual relations among adjacent temporal slices. Therefore, abundant spatio-temporal information can be derived from A and be processed for further OA modeling.

3.2. Dcgan

GANs, with their powerful convolutional architecture, have been successfully applied in data-hungry scenarios. Typically, a DCGAN is composed of two sub-networks: a generative network (denoted by G for generator), and a discriminative network (denoted by D for discriminator). The generator G seeks to synthesize samples by transforming a noise vector z as $x = G(z)$. The discriminator in turn strives to distinguish the generated samples from a latent distribution $p_{\rm fake}(z)$ rather than the real data distribution $p_{\rm real}(x)$. The competition between generator G and discriminator D can be defined as a min–max game:

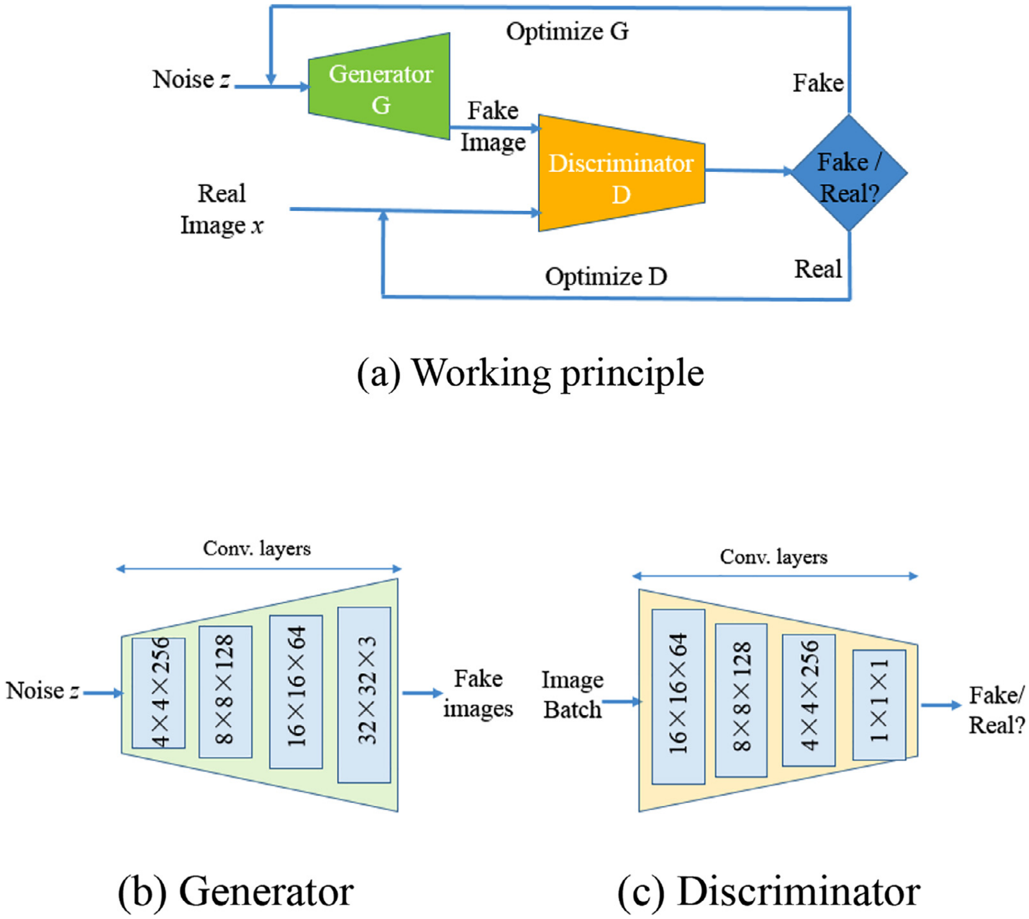
$$\min_G \max_D \mathbb{L}(G,D) = \mathbb{E}_{\rm x\tilde{p}_{real}(x)}[\rm logD(x)] + \mathbb{E}_{\rm z\tilde{p}_{fake}(z)}[\rm log(1 - D(G(z)))] \tag{2}$$

where $D(x)$ and $D(G(z))$ are the probabilities that the sample is from the real data distribution $p_{\rm real}(x)$ and the latent distribution $p_{\rm fake}(z)$, respectively. In the ideal case, the generator G constantly challenges the discriminator D by minimizing the loss function $\mathbb{E}_{\rm z\tilde{p}_{fake}(z)}[\rm log(1 - D(G(z)))]$, generating an increasing number of samples discriminated as real samples; while the discriminator D should be trained and updated jointly by minimizing the loss function $\mathbb{E}_{\rm x\tilde{p}_{real}(x)}[\rm logD(x)]$, classifying the real samples as real and the generative samples as fake.

We illustrate the working principle and model architecture of the DCGAN in Fig. 1. The optimization signal for G is provided by D: if D is able to evaluate the authenticity of candidate images with stride convolutions, *i.e.*, $D(G(z))\approx 0$ and $D(x)\approx 1$, then G will be optimized through fractionally-strided convolutions to produce more realistic samples, and vice versa. As the tug-of-war continues, G competes with D on the distance between $p_{\rm fake}(x)$ and $p_{\rm real}(x)$ as:

$$\begin{aligned} \min_G \max_D \mathbb{L}(G,D) &= \mathbb{E}_{\rm x\tilde{p}_{real}(x)}\Big[\log\frac{\rm p_{real}(x)}{\rm p_{real}(x)+p_{fake}(x)}\Big] + \mathbb{E}_{\rm x\tilde{p}_{fake}(x)}\Big[\log\frac{\rm p_{fake}(x)}{\rm p_{real}(x)+p_{fake}(x)}\Big] \\ &= -2\log2 + \mathrm{KL}\Big[p_{real}(x)||\frac{p_{real}(x)+p_{fake}(x)}{2}\Big] + \mathrm{KL}\Big[p_{fake}(x)||\frac{p_{real}(x)+p_{fake}(x)}{2}\Big] \\ &= -2\log2 + 2\mathrm{JSD}[p_{real}(x)||p_{fake}(x)] \end{aligned} \tag{3}$$

which amounts to optimizing the Jensen-Shannon divergence (JSD). Only when $p_{\rm fake}(x)$ produced by $G(x)$ is similar to $p_{\rm real}(x)$



(a) Working principle

(b) Generator

(c) Discriminator

Fig. 1. The working principle and model architecture of a DCGAN.

and even the optimal *D* cannot tell the differences between two distributions, *G* and *D* converge to a Nash equilibrium [38].

4. Methodology

Recent innovations in a wide range of applications have demonstrated the feasibility of DCGANs, however, they are still far from perfect. First, due to the unsupervised training manner, there exists no constraint on the modes of the samples being created. Second, the stacked convolutions can capture only the multi-scale local dependencies with small receptive fields, while modeling spatial–temporal OA dynamics requires the perception of the global content. Third, it is arduous for a DCGAN to stabilize the training on low-resolution signals. As the gradient decreases during back-propagation, the deep model may encounter a vanishing gradient and fail to adjust the neuron parameters accordingly.

In this section, we present a conditional self-attention GAN framework, named OA-GAN, to overcome above limitations. Specifically, we retrofit the conventional DCGAN and propose efficient but simple modifications to the backbone architecture as follows.

4.1. Incorporating external information for conditional output

Occupant-driven building services call for precise OA modeling, however, generating specified types of samples from multi-activity data is impractical for an unconditioned GAN. For example, if the generative task is to imitate “sitting down”, unconditioned GAN would learn from all input data without direction (*i.e.*, multi-class activity samples), and generate a batch of candidate activity samples which may be “sitting down”, “squatting”, or “arm waving”. All sample labels will be processed using one-hot encoding in the offline stage. To avoid brute-force generation, we attempt to engage both *G* and *D* in different modes by incorporating contextual information. Given an external class label *y* which belongs to an embedding space *Y* drawn from training data, we can define a conditional *G* and *D* along with an embedding $y \sim p_Y(y)$, and replace $p_{\text{real}}(x)$ and $p_{\text{fake}}(z)$ with joint distributions $p_{\text{real}}(x,\ y)$ and $p_{\text{fake}}(z,\ y)$, respectively [39]. Therefore, the task of *G* is to combine the prior latent distribution $p_{\text{fake}}(z)$ with the conditional density $p_Y(y)$ for joint hidden representation, while the task of *D* is to predict the probability that × falls into the real category under condition *y*. The objective function of the min–max competition can be rephrased as follows:

$$\min_G \max_D \mathbb{L}(G,D) = \mathbb{E}_{\mathrm{x,y}\tilde{\mathrm{p}}_{\mathrm{real}}(\mathrm{x,y})}[\mathrm{logD}(\mathrm{x,y})] + \mathbb{E}_{\mathrm{y}\tilde{\mathrm{p}}_{\mathrm{Y}}(\mathrm{y}),\mathrm{z}\tilde{\mathrm{p}}_{\mathrm{fake}}(\mathrm{z})}[\mathrm{log}(1 - \mathrm{D}(\mathrm{G}(\mathrm{z,y}),\mathrm{y}))] \tag{4}$$

In Fig. 2, we illustrate this simple but efficient modification to the working principles of DCGANs by embedding auxiliary label information. Suppose that *G* takes z_i from a Gaussian distribution and feeds a batch $\{(z_i,y_i)\}_{i=1}^{n}$ of conditional samples into *D* with real images $\{(x_i,y_i)\}_{i=1}^{n}$. The cost function of *D* can be simplified as a logistic cost summation for fast convergence:

$$\mathbb{L}_D = -\frac{1}{2n}[\sum_{i=1}^{n} logD(x_i,y_i) + \sum_{i=1}^{n} log(1 - D(G(z_i,y_i),y_i))] \tag{5}$$

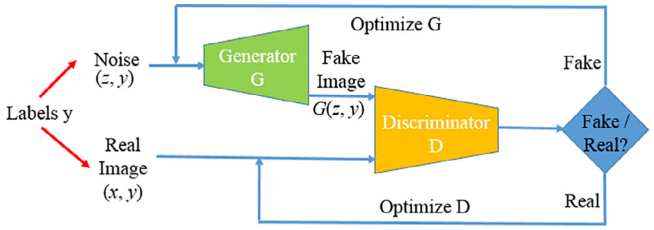


Fig. 2. The working principle of OA-GAN.

If D is able to assign correct labels to the samples, *e.g.*, 1 for (x_i,y_i) and 0 for (z_i,y_i), G will then follow up by leveraging label y_i to direct the training process and maximize the probability that a positive label is assigned to (z_i,y_i):

$$\mathbb{L}_G = -\frac{1}{n}\sum_{i=1}^{n} logD(G(z_i, y_i)) \tag{6}$$

The competition continues until D consistently outputs an approximate probability of 0.5 for both real samples and generated samples, which means D is maximally fooled and G generates radio images within certain categories. Note that the class label y plays an essential part in controlling the learning process. The discriminator D would be expected to first ask for the use of the label y, in that some activity attributes specified in y will be of benefit to minimizing training loss further than an unconditioned GAN might achieve. The generator G would then follow up by leveraging the label y soon after the D learns the proper weights for accepting y.

4.2. Manipulating global dependencies using self-attention mechanism

Recall that we have constructed multi-dimensional radio images for comprehensive analysis, however, it remains a challenge for GAN-based frameworks, *i.e.*, how to take advantage of spatial–temporal dependencies jointly for high-quality image generation and discrimination. Traditional convolutional or recurrent operations can process only spatial or temporal messages from radio images from a local perspective, providing insufficient OA clues for further modeling. Repeating local operations through measurement or combining a CNN-RNN hybrid model can yield a larger reception field, however, they sacrifice the computational efficiency and increase the optimization difficulties. In this paper, we retrofit the performance of convolutions for joint spatial–temporal modeling by means of an emerging self-attention mechanism [29]. In contrast to vanilla convolutions, which focus on adjacent regions in feature maps (millisecond-level variations induced by a few subcarriers), self-attention aims to establish a one-to-one association between any two elements in a radio feature map, and jointly capture the short/long-range dependencies regardless of their space–time distances. Therefore, the self-attention module is able to amplify the implicit correlations behind contextual OA dynamics and suppress the disturbances of unrelated noise.

For clarity, we present the diagram of the self-attention module employed in OA-GAN in Fig. 3. Suppose there are already convolution operations that can incorporate time-varied CSI from multiple subcarriers and output feature maps containing transitory information from adjacent subcarriers. First, we project the radio feature maps $F \in \mathbb{R}^{C\times H\times W}$ into three latent feature spaces, *i.e.*, *Query*, *Key* and *Value*, through a series of 1×1 convolutions. This is because an attention function can be described as mapping a Query and a set of Key-Value pairs to an output, just like “soft addressing” does: when there is a Query request, the corresponding attention value can be retrieved by comparting the similarity of the address of Query and the element Key in the memory.

Specifically, these projections are initially randomized from the same source and reflect all the internal responses in each local position, which can be formulated as $Query(F) = w_qF$, $Key(F) = w_kF$ and $Value(F) = w_vF$, respectively, where $w_q \in \mathbb{R}^{C\times C/4}$,$w_k \in \mathbb{R}^{C\times C/4}$ and $w_v \in \mathbb{R}^{C\times C}$ are the corresponding weight matrices to be learned. By multiplying F by learnable matrices w, we obtain a better representation for measuring the compatibility between elements. Here, C, H, and W are the channel number, height and width, respectively, of the radio feature maps, and $(H\times W)$ is the total number of elements in the current feature map.

Given a candidate $Query(F_b)$ and a set of target $Key(F_a)/Value(F_a)$ pairs, then we calculate the attention weights by measuring the similarity value S among the internal elements. Note that we circumvent the use of additional neural networks to simplify the process and instead calculate the dot product between *Query* and each *Key* as $S_{a,b} = Query(F_b).Key(F_a)^T$, where a and $b \in \mathbb{R}^{H\times W}$. Thus, S can suggest the contribution of each element in *Query* by computing a weighted sum of the responses from all possible regions in *Key*. In addition, we adopt the softmax function as an image mask to normalize the attention value as $\alpha_{b,\ a}$ and multiply it by $Value(F_a)$ to obtain the weighted attention map:

$$atten_b = \sum_{a=1}^{H\times W} Value(F_a).\alpha_{b,a} \tag{7}$$

where $\alpha_{b,a} = e^{s_{b,a}} / \sum_{a=1}^{H\times W} e^{s_{b,a}}$denotes the proportion of the a-th attention paid to the b-th input. The weighted attention map is further adjusted by a learnable parameter γ and then added back to the original input for refined feature maps:

$$O = \gamma \times atten + F \tag{8}$$

Therefore, attention-refined feature maps can gradually understand the whole content by assigning higher weights to context-related regions and lower weights to isolated elements. Since the self-attention module requires no external supervisions and sophisticated transformations, it is compatible and lightweight for any existing deep learning backbones to integrate into. In practical applications, we implement self-attention modules in the early stages of radio image generation and discrimination.

4.3. Constructing a deeper model using residual connections

Recall that our ultimate goal is to train a generator G that produces equally good radio images accepted by a discriminator D, and G constantly competes with D until D is maximally fooled. However, D can easily dominate the competition as G fails to learn distinct representations from low-resolution radio images, and G receives no feedback from D and loses incentives to update its parameters as the gradients of D approach zero at almost every iteration [41]. Therefore, we seek to strengthen the generation capability by establishing a deeper model for generator G. However, this process may also aggravate the problem of vanishing gradients.

Motivated by works on super-resolution images [42], we propose a deep residual generator and a relatively shallow discriminator for efficient competition, whose architectures are illustrated in Fig. 4(a) and 4(b), respectively. Specifically, to train a proper discriminator D that can deliver feedback and drive G to update the parameters, we follow the guidelines in [36] and design a cascade module for attention-refined feature maps in D. This module consists of 6 stacked convolutional (Conv.) layers, and each of them is followed by a leaky rectified linear unit (LeakyReLu) layer, a batch normalization (BN) layer and a dropout (Drop) layer, consequently. Here, we increase the kernel number (N) layer by layer in order to deepen the understanding of the multi-scale variations, while the sizes of both kernels (K) and strides (S) are fixed at 3×3 and

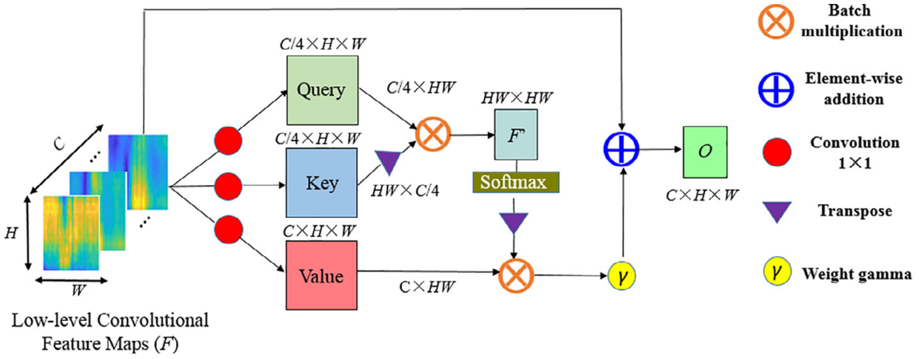


Fig. 3. Diagram of the self-attention module employed in OA-GAN.

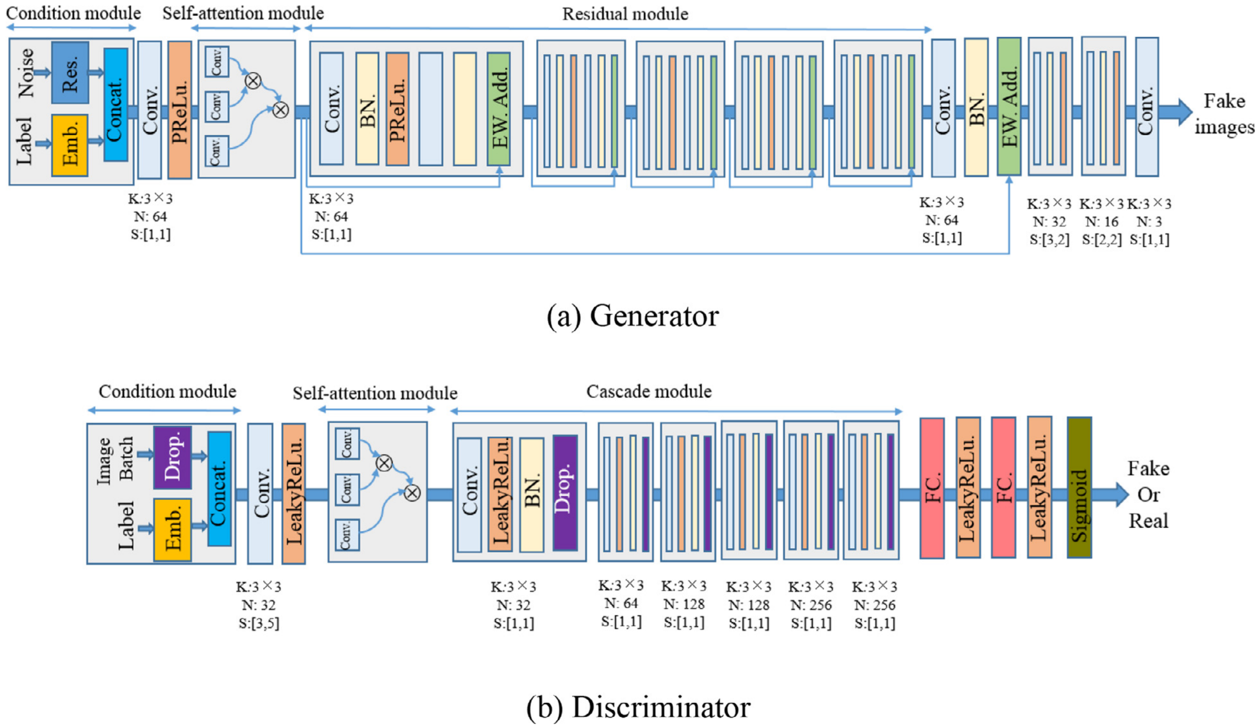


Fig. 4. The architecture of the (a) generator and (b) discriminator in the OA-GAN framework.

[1,1]. The LeakyRelu layer performs a simple activation that multiplies any negative input values by scale value of 0.2; the BN layer controls the sensitivity towards variations within the samples across a min-batch; and the Drop layer is used to prevent overfitting by randomly setting the sample values to zero with a low probability of 0.25. The abstract feature maps are then fed into two fully-connected (FC) layers for dense mapping, and a sigmoid layer outputs the probability from 0 to 1 for final evaluation.

To train a powerful generator that produces vivid radio images, we resort to a novel generation strategy. The core of the innovation lies in the construction of a deep residual module, which serves two purposes. On the one hand, it ensures unhindered gradient flow by adopting simple residual connections, and further increases the depth (rather than the width) of the network for more comprehensive feature learning. On the other hand, since shallower features (*i.e.*, features that are close to the input layer) reflect the local context but are general to the output categories, whereas deeper features (*i.e.*, features that are close to the output layer) reveal the global content but are specific to the classification tasks, both local information and global information can be integrated through skip connections for informed output [43]. To this end, we design a deep residual module that connects the shallow layers and deep layers. Empirically, the module includes 5 identical residual blocks, each of which consists of two 3×3 transposed Conv layers with a stride of 1, followed by two BN layers and a parametric ReLu (PReLu) layer as the activation function. To fit it back into the former input, we crop the feature maps to the same size. The resulting 64 feature maps containing both the local and global clues are then processed by three additional transposed Conv layers to recover its resolution, and the final output is the same size as the realistic radio images.

5. Experimental evaluation

In this section, we first present the experimental settings, including the dataset description and training details. Then, we demonstrate the superiority of the OA-GAN by visualizing the outputs and comparing the image similarities with baselines, and conducting diverse GAN-based experiments on common classification methods.

5.1. Experimental settings

5.1.1. Dataset description

In this work, we conducted experiments on the WiFi-based Activity Recognition (WiAR) dataset, an up-to-date public dataset proposed in [44], in consideration of three aspects. First, the WiAR dataset includes various types of activities commonly performed in occupants' daily lives, and thus we can evaluate the performance of multi-class generation and provide optimized solutions for demand-driven services. Second, the WiAR dataset collects thousands of CSI instances from different volunteers in diverse indoor scenarios to ensure sample diversity for robustness evaluation. Third, the WiAR dataset constructs the datasets with the fewest number of WiFi devices, increasing the difficulties in fine-grained OA modeling.

Specifically, the WiAR dataset contains 16 activities performed by 10 users (5 males and 5 females) in 3 indoor environments (*i.e.*, an empty room, a meeting room and an office). All CSI instances are collected by two laptops equipped with Intel 5300 NICs. One laptop with three external antennas acted as the receiver and another laptop with one external antenna acts as the transmitter. Here, the 802.11n CSI Tool [40] provides 30 subcarriers for each antenna pair and the transmitting rate is set to 30 packets per second. During data collection, one volunteer was located in the middle of the propagation path and each activity was repeated 30 times. Each sample was collected by manually starting and stopping under the instruction of a laboratory staff member, resulting in diverse measurement lengths. To maintain the same size, we augmented the initial measurements by adding Gaussian noises to both ends. As a result, the final CSI measurements have a unified shape of 30 × 500 × 3, and we normalize the input values to the range of [-1, 1] for further processing.

5.1.2. Training details

All the models evaluated in this paper were processed with MATLAB 2020a on a Windows 10 desktop equipped with an Intel i7-8700 K processor, 11 GB memory, and an NVIDIA GTX 1080 GPU. At the early training stage, as discriminator D converged faster than the generator G, we set the training rates of D and G to 10^{-4} and 5×10^{-5}, respectively, to balance the competition. In generator G, a noise prior z with the dimension of 100 was projected into a shape of 5 × 125 × 512 and then concatenated with label y, followed by a Conv. layer with the activation of PReLU, with layer sizes of 64 and 128, respectively, before both being mapped into the self-attention module and the residual module for generating the 30 × 500 × 3 fake images. In discriminator D, real images × were randomly dropped out with a probability of 0.25, and the labeled images (x, y) were then mapped into a Conv. layer with the activation of LeakyReLu, with layer sizes of 32 and a scale of 0.2 respectively, before being fed into the subsequent modules for discrimination. To train the models, an adaptive moment estimation (Adam) optimizer was used to update the parameters with a gradient decay factor of 0.5. The maximum number of epochs was set to 50, and each mini-batch with 64 observations was used for gradient descent. We randomly flipped the labels with a probability of 0.2 for stable generator G and utilized one-sided label smoothing with a value of 0.9 in discriminator D for stable training.

To monitor how well G and D achieve their respective goals, we plot their real-time scores in Fig. 5(a) and 5(b). The G score is the average probability corresponding to the output of D for the generated images, while the D score is the average probability of the input images belonging to the correct class. We notice that both scores fluctuate significantly in the early training stage (D dominates G), and D reaches a steady stage of approximately 0.5 as the competition continues, which indicates that G and D achieve a harmonious trade-off. To further diagnose issues during training, we also visualize the generator output (500 × 480 × 3) on the 250th, 700th, 1500th and 2500th iterations, where the x-axis and y-axis denote the packet number (total of 500 packets) and activity index (total of 16 activities, each one with 30 subcarriers), respectively. As shown in Fig. 5(c)-(f), we observe that G can learn the rich feature representations for latent vector projection and gradually enrich the fine-grained information in the generated images, which demonstrates the effectiveness of the competition design.

5.2. Experimental results

In this section, we illustrate the experimental results by visualizing the synthetic radio images and compare the generation similarity of OA-GAN with other SOTA frameworks.

5.2.1. Visualized results

To present our general intuition on the generation performance, we first visualize the virtual 3-channel CSI radio images generated by the proposed OA-GAN framework. Each channel contains the synthesized amplitude information belonging to each receiving antenna. Due to the layout limitation, we display only the generation results on 6 challenging activities in Fig. 6, including 2 upper-body activities (horizontal arm wave and high arm wave), 2 lower-body activities (forward kick and side kick) and 2 whole-body activities (squat and sit down). From Fig. 6, we discover two essential observations. 1) Each antenna can produce similar and informative subcarrier-level variations for the same spatial–temporal event. For example, when G models the "horizontal arm wave" activity, all 30 subcarriers for each antenna can induce corresponding fluctuations regarding both the spatial and temporal relations, and reveal the movement patterns in each channel. 2) Even similar activities can lead to an obvious discrepancy in the synthesized signal patterns of diverse antennas and further result in unique radio images. For example, the "forward kick" and "side kick" activities are semantically quietly similar except for the motion direction, while G can still identify slight differences and produce distinct radio images for sample diversity. The above evidence strongly suggests that with the efficient network design, the generator G not generates only the representative samples for each activity but also avoids model collapse, which replicates only a small variety of signal patterns.

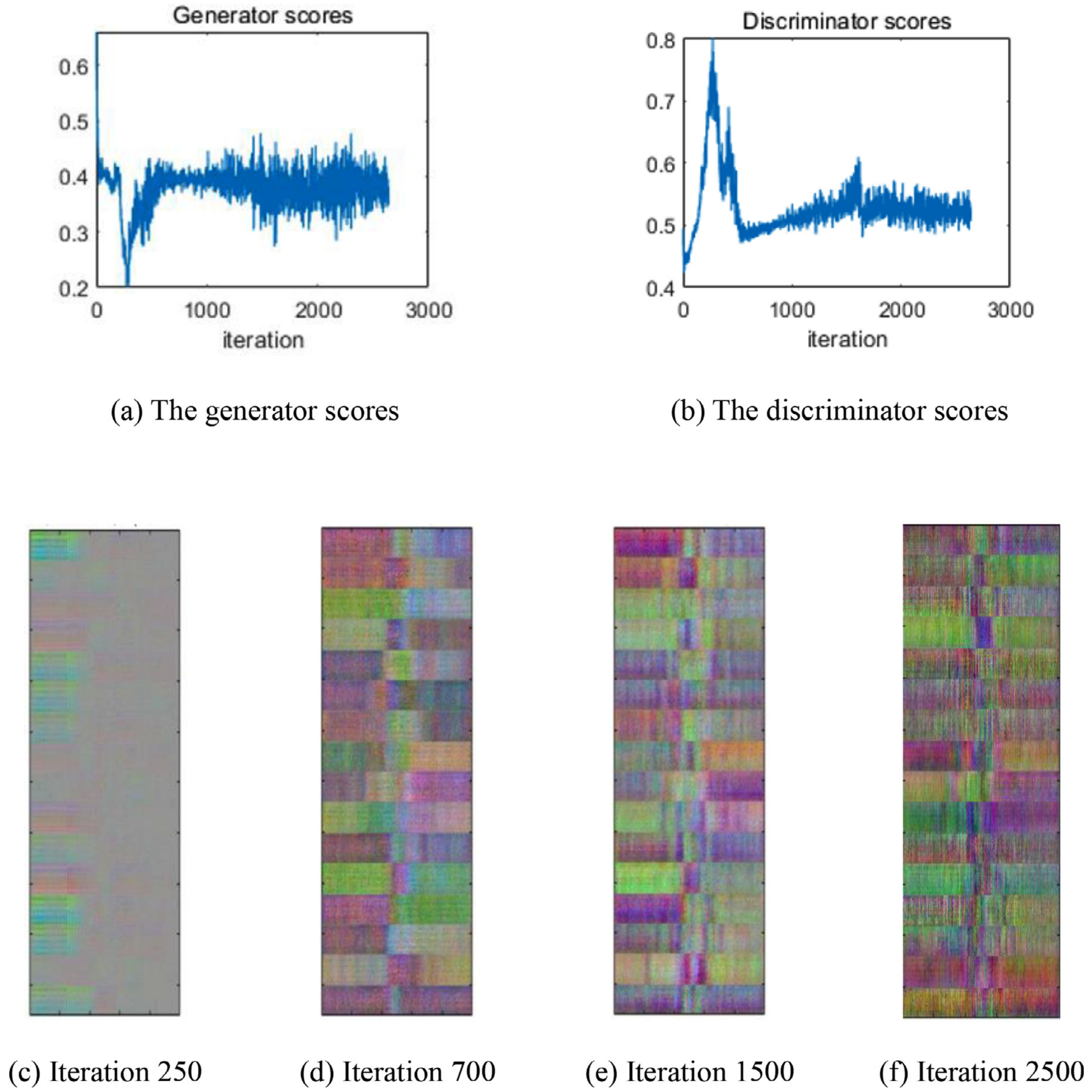
5.2.2. Comparison of image similarity

To evaluate the quality of the generated images, we measure the structural similarities (SSIM) index between the real images and synthesized images [45]. Ideally, an OA model should be able to synthesize the radio images that share similar semantic clues with ground-truth measurements. However, measuring the differences directly is not trivial because the statistical properties (*e.g.*, color, texture and shape) are obscured. Therefore, we consider the visual saliency of the radio images and select a hybrid SSIM to indicate the latent similarity regarding image luminance, contrast and structure. Specifically, the luminance L, contrast C and structure S between real image $I_{x,\,y}$ and generated image $I_{z,\,y}$ can be defined as:

$$L(I_{x,y}, I_{z,y}) = \frac{2\mu_{x,y}\mu_{z,y} + u_1}{\mu_{x,y}^2 + \mu_{z,y}^2 + u_1} \tag{9}$$

$$C(I_{x,y}, I_{z,y}) = \frac{2\sigma(I_{x,y} - \mu_{x,y})\sigma(I_{z,y} - \mu_{z,y}) + u_2}{\sigma^2(I_{x,y} - \mu_{x,y}) + \sigma^2(I_{z,y} - \mu_{z,y}) + u_2} \tag{10}$$

$$S(I_{x,y}, I_{z,y}) = \frac{\theta\left(\frac{I_{x,y} - \mu_{x,y}}{\sigma_{x,y}}, \frac{I_{z,y} - \mu_{z,y}}{\sigma_{z,y}}\right) + u_3}{\sigma(I_{x,y} - \mu_{x,y})\sigma(I_{z,y} - \mu_{z,y}) + u_3} \tag{11}$$



(a) The generator scores

(b) The discriminator scores

(c) Iteration 250

(d) Iteration 700

(e) Iteration 1500

(f) Iteration 2500

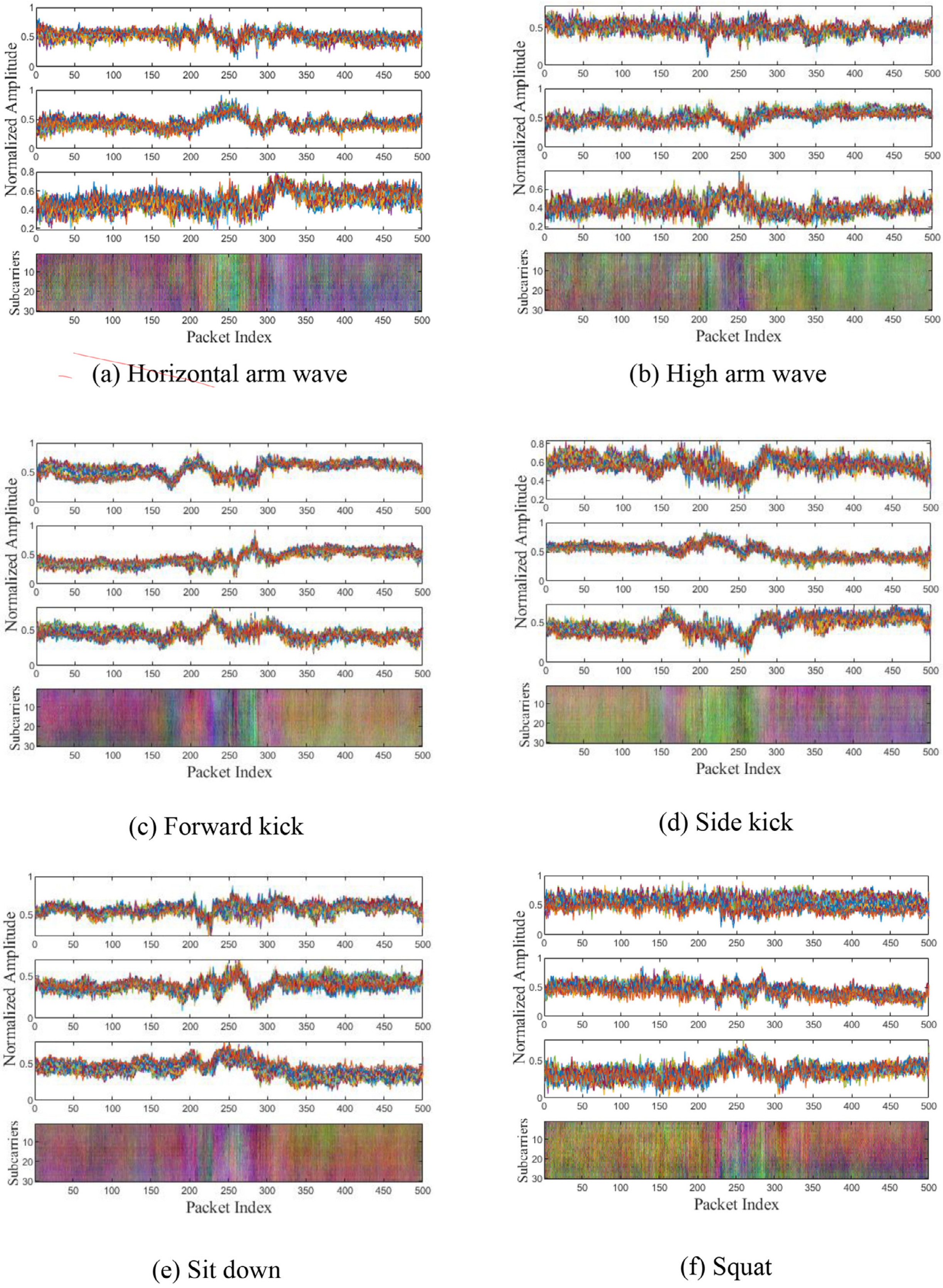
Fig. 5. The training progress of the proposed OA-GAN framework, where (a) and (b) is the real-time scores of the generator and discriminator respectively, and (c)-(f) are the observations corresponding to various generation stages.

where $\{\mu, \sigma, \theta\}$ denotes the mean, standard deviations and covariance of the target radio images, respectively. Similar constants $\{u_1, u_2, u_3\}$ are applied in formulas to avoid dividing by zero. We formulate the SSIM as a combination of the above power-exponential functions:

$$\mathrm{SSIM}(I_{x,y}, I_{z,y}) = L(I_{x,y}, I_{z,y})^{\alpha} * C(I_{x,y}, I_{z,y})^{\beta} * S(I_{x,y}, I_{z,y})^{\gamma} \quad (12)$$

and the SSIM score for each generated image $I_{z,\ y}$ can be obtained by comparing the generated images with each real image $I_{x,\ y}$ in a specific class and computing the average values. Higher SSIM scores suggest higher similarity between the real images and the generated images and further indicate better generation quality.

We have also conducted generation tasks on three SOTA approaches (with conditional labels) to quantify the image similarity. 1) Building-GAN [17]. A pioneering work proposed by Chen et al that uses vanilla GAN to model occupancy with camera data. We modify the FC architecture to take the CSI amplitudes in. 2) Amplitude feature DCGAN (AF-DCGAN) [35]. The first work that applies CSI amplitude variations to occupant localization based on a DCGAN. The implementation of this model adopts the same convolution settings as OA-GAN. 3) Super-resolution GAN (SRGAN) [42]. A popular GAN-based framework that recovers photo-realistic textures from down-sampled photos using a deep residual GAN. We adjust the parameters as well as the connection settings to fit the process of CSI radio images. Table 1 illustrates the SSIM scores obtained by the four GAN-based approaches for 16 activities, and all SSIM scores are average values comparing the similarity between all generated samples and real samples. We can see that all the GAN-based approaches obtain relatively higher SSIM scores on the whole-body activities (the bend, walk, sit down and squat activities), while suffering performance degradation when modeling partial-body movements (especially for the forward kick and side kick activities) to varying degrees. Specifically, the maximum SSIM score for any activity is 0.733 achieved by OA-GAN, indicating the essential features of activity “walk” are generally captured and the generated samples are comparatively realistic. While the minimum SSIM score appears in modeling “side



(a) Horizontal arm wave

(b) High arm wave

(c) Forward kick

(d) Side kick

(e) Sit down

(f) Squat

Fig. 6. Visualization of the synthetic amplitudes for each antenna and the corresponding radio images for occupant activities: (a) horizontal arm wave, (b) high arm wave, (c) forward kick, (d) side kick, (e) sit down and (f) squat.

Table 1
The SSIM scores for the four GAN-based approaches on the WiAR dataset.

Methods/Activity	Building-GAN [17]	AF-DCGAN [35]	SRGAN [42]	The proposed OA-GAN
Horizontal arm wave	0.406	0.495	0.582	0.655
High arm wave	0.418	0.487	0.608	0.651
Two hands wave	0.421	0.504	0.619	0.673
High throw	0.393	0.469	0.597	0.659
Draw an “X”	0.413	0.495	0.588	0.643
Draw a tick	0.401	0.488	0.591	0.658
Toss paper	0.425	0.509	0.606	0.662
Forward kick	0.381	0.443	0.546	0.637
Side kick	0.377	0.451	0.552	0.641
Bend	0.497	0.584	0.688	0.726
Hand clap	0.433	0.516	0.624	0.687
Walk	0.501	0.612	0.691	0.733
Phone call	0.432	0.511	0.617	0.704
Drink water	0.424	0.523	0.605	0.698
Sit down	0.508	0.596	0.681	0.732
Squat	0.515	0.593	0.675	0.729

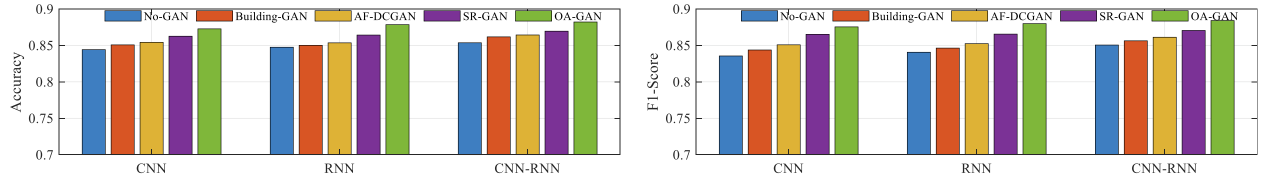
kick”, which is only 0.377 provided by Building-GAN due to ambiguous motion patterns. The reason is because partial-body movements affect the propagation paths of WiFi signals in an unapparent manner, and thus induce non-significant variations for identification. The convolutional operations can help AF-DCGAN to model these implicit relations, and produce a 0.083 higher average SSIM score. The efficient design of residual convolution blocks can further help SRGAN and OA-GAN to increase the model depth and enhance the generation quality by a clear margin. In addition, we also notice that obvious leaps occurred in modeling complex spatial–temporal events. By employing the efficient self-attention module, OA-GAN surpasses SRGAN in each activity category, and achieves a relatively higher improvement especially for the “phone call” and “drink water” activities, which verifies the capability of spatial–temporal modeling.

5.2.3. Comparison of the recognition performance

Since the ultimate goal of OA-GAN is to augment the sample diversity and increase the OAR performance, it is insufficient to judge the generation results using only the SSIM index. The reason is because generator *G* may fail to learn a rich representation from the CSI measurements and further duplicate identical images for diverse OA patterns, which does little to help enrich the dataset. To fully consider the sample diversity, we apply two different generation strategies to two data-hungry scenarios using the above GAN-based approaches. In the 1st scenario where volunteers perform each activity only 20 times, we construct a training set with a total of 20 (times) × 16 (activities) × 10 (volunteers) = 3200 instances (leaving the remaining 1600 instances as the testing set), and generate 200 radio images (similar to an additional10 volunteers performing 20 times) for each activity category. For the 2nd scenario with only 7 volunteers engaged in, we construct another training set with a total of 30 (times) × 16 (activities) × 7 (volunteers) = 3360 instances (leaving the remaining 1440 instances as the testing set), and generate 210 radio images (similar to 7 visual volunteers performing the activities30 times) for each activity category. To fairly quantify the contribution, we feed the training sets into three neural-network-based classifiers, *i.e.*, CNN [24], RNN [20] and a hybrid CNN-RNN [13] model, with proper parameter settings respectively, and use the corresponding testing sets to validate the performance of the GAN-based approaches. Two common metrics called the “recognition accuracy” and “F1-score” are leveraged for evaluation. Here, the accuracy refers to the ratio of OA patterns classified with correct class labels to the total number of labels, and the F1-score is the harmonic mean of the precision (the proportion of positive instances identified by classifier) and recall value (the proportion of positive samples in the original dataset).

The evaluation results of the GAN-based approaches under diverse data-hungry scenarios are shown in Fig. 7, and Fig. 8. From Fig. 7, we can discover that all the classifiers can benefit from the generation results and attain satisfactory performance, which is particularly crucial for the scenario where users are unwilling to repeat an activity several times. In such cases, GANs are able to learn invariant representations from a few samples and emulate the users’ performing habits. Moreover, we also notice that OA-GAN always outperforms the other baselines by a small margin. This finding indicates that although neural-network-based classifiers have been specifically modified, OA-GAN can still contribute to the recognition results by providing virtual motion clues, especially for spatial–temporal manipulations. Similar observations can be obtained from Fig. 8. In the 2nd scenario where only 7 users (labels 1–7 in the dataset) participate in the collection task and 3 users (labels 8–10) are left out for validation, all classifiers witness the performance degradation as the new-coming OA instances may contradict with the patterns to be trained. By integrating a deeper convolutional architecture with residual connections and a self-attention mechanism for long-term dependencies, OA-GAN offers promising candidates for unknown users, and drives all the classifiers to perform much better than the condition without the implementation of the GAN. This finding suggests that GAN-based approaches can help classifiers to overcome the bottleneck of identifying new types of OAs with similar semantic meanings. In addition, compared with other SOTA baselines, OA-GAN shows its superiority by achieving the largest increment of accuracy and F1-score for all three classifiers. This finding implies the powerful capability of OA-GAN to retrieve the fine-grained radio images with distinct local–global relations.

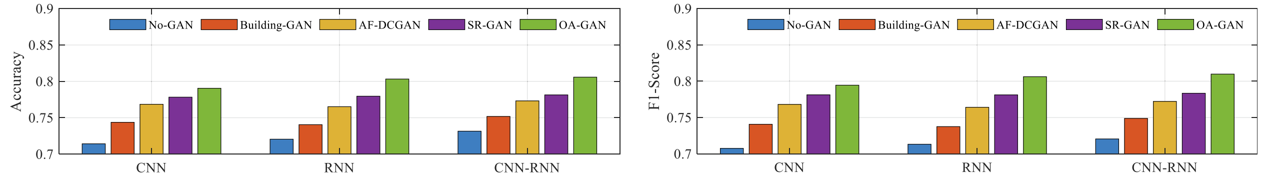
Additionally, we also evaluate the impact caused by the number of virtual training samples generated by OA-GAN in both scenarios. From Fig. 9 and Fig. 10, we observe that as more number virtual samples are added to the training set, all the classifiers attain important gains in handling insufficient datasets, especially for the 2nd scenario where the gap of individual diversity remains large, which confirms the effectiveness of the proposed OA-GAN framework. Moreover, we can also witness a similar trend in improvement, where the rate is rather high at the early stage and then slows down at the final stage. This finding reveals that OA-GAN can synthesize a batch of high-quality radio images, yet reach an upper limit as soon as the generator fails to produce more efficient dynamics from limited sources. In the future, we will seek to transfer the abundant knowledge from semantically similar domains and prompt the generation capability using multi-modal data sources, such as audio and video records.



(a) Accuracy

(b) F1-Score

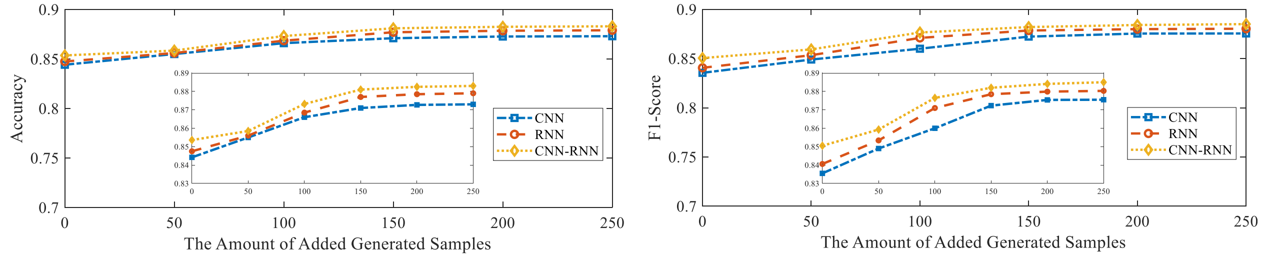
Fig. 7. Performance comparison in the 1st scenario where volunteers perform the activities in each activity category fewer times.



(a) Accuracy

(b) F1-Score

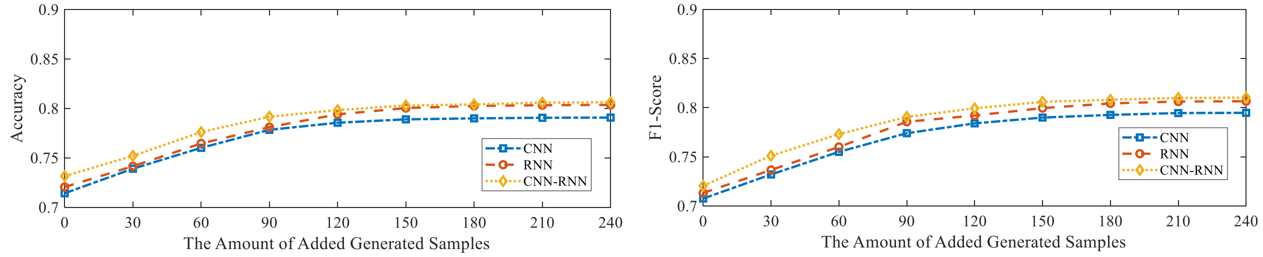
Fig. 8. Performance comparison in the 2nd scenario where fewer volunteers perform the activities in each activity category a normal number of times.



(a) Accuracy

(b) F1-score

Fig. 9. The performance changes as the amount of added virtual samples increases in the 1st scenario where volunteers perform each activity fewer times.



(a) Accuracy

(b) F1-score

Fig. 10. The performance changes as the amount of added virtual samples increases in the 2nd scenario where fewer volunteers perform the activities.

6. Discussion and limitations

6.1. Complex activity modeling with environmental interferences

Since WiFi sensing technology is still in its infancy, the scale of existing data sets is generally small, and the types of activities included are not rich (usually short-term activities). Under the conditions of limited but simple OA samples, OA-GAN effectively augments the existing datasets, including single-limb movement modeling (*e.g.*, high arm wave) and compound-limb movement modeling (*e.g.*, walk), which provides a possible solution to the problem of data hunger. In addition, OA-GAN can be easily applied for a large amount of short-term activities through simple model adjustment. There are two reasons: First, OA-GAN can characterize the discriminative features of active signals from the local and global perspectives, thereby avoiding interference from similar activities. Second, previous works (*e.g.*, SignFi [24]) have verified that only adopting the 5-layer CNN model can effectively distinguish 276 standard gestures through WiFi. Based on this, OA-GAN further adopts the residual connections for the convolution module, which can ensure the effective transmission of training gradients and continuous update of training parameters.

However, in the real scenarios, OA patterns are more complex (including both long-term and short-term activities) and OA-related signals easily affected by a variety of environmental factors, such as different indoor layouts, occupant locations and orientations. To discover the intrinsic patterns from complex OAs, OA-GAN may require more training samples to cover the impacts of multiple variables and thus overcomes the influences of environmental dynamics using more computing resources. In the future work, we will study how to separate effective OA information from entangled WiFi signals, thereby further reducing the requirement on the training samples.

6.2. Similar activity modeling with fewer OA samples

By visualizing the experimental results in Fig. 6 and quantifying the experimental results in Tab 1, it is shown that OA-GAN can learn the differential characteristics of similar activities after obtaining a certain number of samples, *e.g.*, "Horizontal arm wave" vs "High arm wave", "Forward kick" vs "Side kick", "Sit down" vs "Squat". Further, we attempt to minimize the number of training samples to discover the ability of OA-GAN to cope with extreme situations of data scarcity. Specifically, OA-GAN is trained with only one sample from each user per activity and is guided to generate 4 candidate images for each activity, which is shown in Fig. 11. It is clearly observed that OA-GAN can hardly generate discriminative activity samples. We summarize the reasons for the failure as follows: On the one hand, OA-GAN is a kind of deep neural network, the lack of OA samples can cause overfitting of model training. On the other hand, due to severe environmental dynamics, it is almost impossible for OA-GAN to directly learn the intrinsic laws from inconsistent OA-related signal patterns using only a small number of training samples. In the future work, we will conduct an in-depth study related work based on zero-shot learning, so as to achieve stable generation of small or even missing samples.

6.3. Multi-person activity modeling with commercial WiFi devices

In real building scenarios, users usually deploy only a pair of WiFi transceiver devices in one room. However, due to inherent deficiencies of commercial WiFi devices, such as poor transmission quality, limited communication bandwidth and number of antennas, it is non-trivial for WiFi-based technology to support the sensing task of multi-person activities at current stage. Therefore, this paper mainly focuses on single-person activity modeling based on existing WiFi infrastructure. With the increasing popularity of low-cost wireless devices, we will attempt to modify the structure of OA-GAN in the future so as to leverage multi-sensor information (such as multiple WiFi transceivers, RFID and Bluetooth, etc) for fine-grained multi-person activity modeling.

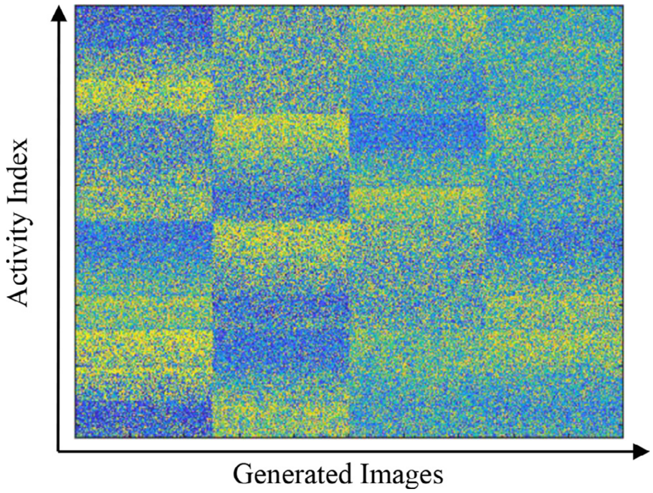


Fig.11. Generated images of 16 activities with only one sample from each user per activity.

6.4. Fixed-length requirement for time-varied signal input

Generally, the Generator/Discriminator requires the input signal to maintain a fixed length. Otherwise, the weight parameter cannot be stabilized, which will cause the network to change dynamically and fail to achieve the purpose of parameter training. From WiAR dataset, we have observed that the activity duration is around 10–15 s and the OA-related signal length is time-varied. To fulfill the requirement of model input, we adopt the conventions of domain-related work that unifies the signal length (*i.e.*, 500) by filling in Gaussian noises in the preprocessing stage. However, the duration of OA is quite random in real scenarios, and the above hard-alignment method may induce invalid computation in short-term activities and information loss in long-term activities. In response to the above problems, we have made early attempt on the design of the flexible CNN based on the saliency region selection [19], which can partly alleviate the contradiction between fixed-length requirement and time-varied signal length assisted by manual efforts. We will further introduce the design of flexible input into GAN model and leverage unsupervised attention information to select salient signal snippets, so as to improve the practical value of OA-GAN in the future.

7. Conclusion

In this paper, we investigated the possibility of non-intrusive OA modeling for occupant-centered services and proposed a WiFi CSI-based OA generative framework, called OA-GAN, to synthesize virtual radio images for dataset augmentation. OA-GAN originated from a zero-sum game between two neural networks, and we retrofitted vanilla GAN with simple yet efficient modifications. First, we introduced external constraints for the conditional output. Second, we properly manipulated the local–global dependencies for comprehensive image retrieval. Third, we leveraged the shortcut connections to construct a deeper generator for informative details. Based on the WiAR dataset, we visualized the generation results and compared their SSIM scores with four SOTA baselines. In addition, extensive experiments were conducted on two data-hungry scenarios, and the experimental results further verified the superiority of the proposed OA-GAN framework.

Funding

This work is partially supported by the National Key Research and Development Project of China under grant No. 2017YFC0704100 (entitled New Generation Intelligent Building Platform Techniques).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N.E. Klepeis, W.C. Nelson, W.R. Ott, J.P. Robinson, A.M. Tsang, P. Switzer, J.V. Behar, S.C. Hern, W.H. Engelmann, The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants, J. Expo. Anal. Environ. Epidemiol. 11 (2001) 231–252.

[2] International Energy Agency, F. Birol, World energy outlook 2015, Paris: International Energy Agency, 2015.

[3] M.S. Andargie, M. Touchie, W. O'Brien, A review of factors affecting occupant comfort in multi-unit residential buildings, Build. Environ. 160 (2019) 106182.

[4] T.A. Nguyen, M. Aiello, Energy intelligent buildings based on user activity: a survey, Energy Build. 56 (2013) 244–257.

[5] H. Chen, S.H. Cha, T.W. Kim, A framework for group activity detection and recognition using smartphone sensors and beacons, Build. Environ 158 (2019) 205–216.

[6] T. Hong, D.a. Yan, S. D'Oca, C.-F. Chen, Ten questions concerning occupant behavior in buildings: the big picture, Build. Environ 114 (2017) 518–530.

[7] H. Zou, Y. Zhou, H. Jiang, S.C. Chien, L. Xie, C.J. Spanos, WinLight: a WiFi-based occupancy-driven lighting control system for smart building, Energy Build. 158 (2018) 924–938.

[8] J. Kim, Y. Zhou, S. Schiavon, P. Raftery, G. Brager, Personal comfort models: predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning, Build. Environ. 129 (2018) 96–106.

[9] A. Aryal, A. Ghahramani, B. Becerik-Gerber, Monitoring fatigue in construction workers using physiological measurements, Autom. Construct. 82 (2017) 154–165.

[10] S.H. Cha, J. Seo, S.H. Baek, C. Koo, Towards a well-planned, activity-based work environment: automated recognition of office activities using accelerometers, Build. Environ. 144 (2018) 86–93.

[11] J.W. Dziedzic, Y. Da, V. Novakovic, Indoor occupant behaviour monitoring with the use of a depth registration camera, Build. Environ. 148 (2019) 44–54.

[12] Z. Wang, B. Guo, Z. Yu, X. Zhou, Wi-Fi CSI-based behavior recognition: From signals and actions to activities, IEEE Commun. Mag. 56 (2018) 109–115.

[13] H. Zou, Y. Zhou, J. Yang, C.J. Spanos, Towards occupant activity driven smart buildings via WiFi-enabled IoT devices and deep learning, Energy Build. 177 (2018) 12–22.

[14] C. Wu, F. Zhang, Y. Hu, K.R. Liu, GaitWay: monitoring and recognizing gait speed through the walls, IEEE Trans. Mobile Comput. 1 (2020) 1–15.

[15] Q. Zhou, J. Xing, Q. Yang, Device-free occupant activity recognition in smart offices using intrinsic Wi-Fi components, Build. Environ 172 (2020) 106737, https://doi.org/10.1016/j.buildenv.2020.106737.

[16] L. Zhang, Q. Gao, X. Ma, J. Wang, T. Yang, H. Wang, DeFi: robust training-free device-free wireless localization with WiFi, IEEE Trans. Veh. Technol. 67 (9) (2018) 8822–8831.

[17] Z. Chen, C. Jiang, Building occupancy modeling using generative adversarial network, Energy Build. 174 (2018) 372–379.

[18] S. Yao, S. Hu, Y. Zhao, A. Zhang, T. Abdelzaher, Deepsense: a unified deep learning framework for time-series mobile sensing data processing, in Proc. of the 26th International Conf. on World Wide Web (WWW'17), April 2017 351-360.

[19] Q. Zhou, J. Xing, W. Chen, X. Zhang, Q. Yang, From signal to image: enabling fine-grained gesture recognition with commercial Wi-Fi devices, Sensors 18 (9) (2018) 3142.

[20] Z. Chen, L.e. Zhang, C. Jiang, Z. Cao, W. Cui, WiFi CSI based passive human activity recognition using attention based BLSTM, IEEE Trans. on Mobile Comput. 18 (11) (2019) 2714–2724.

[21] C. Feng, S. Arshad, S. Zhou, D. Cao, Y. Liu, Wi-multi: a three-phase system for multiple human activity recognition with commercial wifi devices, IEEE Internet Things J. 6 (2019) 7293–7304.

[22] C. Xiao, D. Han, Y. Ma, Z. Qin, CsiGAN: robust channel state information-based activity recognition with GANs, IEEE Internet Things J. 6 (2019) 10191–10204.

[23] S. Palipana, D. Rojas, P. Agrawal, D. Pesch, FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices, in Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) 1(4), 2018, 1-25.

[24] Y. Ma, G. Zhou, S. Wang, H. Zhao, W. Jung, SignFi: Sign language recognition using WiFi. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) 2(1), 2018, 1-21.

[25] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, Z. Yang, Zero-effort cross-domain gesture recognition with Wi-Fi, in Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services (Mobisys'19), 2019, 313-325.

[26] B. Sheng, F.u. Xiao, L. Sha, L. Sun, Deep spatial–temporal model based cross-scene action recognition using commodity WiFi, IEEE Internet Things J. 7 (4) (2020) 3592–3601.

[27] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative adversarial nets, in Proceedings of Advances in neural information processing systems (NIPS), 2014, 2672-2680.

[28] J. Wang, Q. Gao, X. Ma, Y. Zhao, Y. Fang, Learning to sense: deep learning for wireless sensing with less training efforts, IEEE Wirel. Commun., 2020.

[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, I. Polosukhin, Attention is all you need, in Advances in neural information processing systems (NIPS), 2017, 5998-6008.

[30] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016, 770-778.

[31] N. Li, G. Calis, B. Becerik-Gerber, Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations, Autom. Construct. 24 (2012) 89–99.

[32] J. Wang, N.C.F. Tse, J.Y.C. Chan, Wi-Fi based occupancy detection in a complex indoor space under discontinuous wireless communication: a robust filtering based on event-triggered updating, Build. Environ. 151 (2019) 228–239.

[33] T. Leephakpreeda, Adaptive occupancy-based lighting control via grey prediction, Build. Environ. 40 (7) (2005) 881–886.

[34] B. Dong, D.a. Yan, Z. Li, Y. Jin, X. Feng, H. Fontenot, Modeling occupancy and behavior for better building design and operation—a critical review, Build. Simul. 11 (5) (2018) 899–921.

[35] Q. Li, H. Qu, Z. Liu, N. Zhou, W. Sun, S. Sigg, J. Li, AF-DCGAN: Amplitude feature deep convolutional GAN for fingerprint construction in indoor localization systems, IEEE Transactions on Emerging Topics in Computational Intelligence, 2019.

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

[37] S. Yu, H. Chen, E. B. Garcia Reyes, N. Poh, Gaitgan: Invariant gait feature extraction using generative adversarial networks, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR), 2017, 30-37.

[38] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local Nash equilibrium, in Advances in neural information processing systems (NeurIPS), 2017, 6626-6637.

[39] J. Gauthier, Conditional generative adversarial nets for convolutional face generation, Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester, (5), 2014: 2.

[40] D. Halperin, W. Hu, A. Sheth, et al., Tool release: gathering 802.11 ntraces with channel state information, in ACM SIGCOMM Comput. Commun. Rev., 41(1), 2011, 53-53, 2011.

[41] M. Wiatrak, S. V. Albrecht, Stabilizing generative adversarial network training: a survey, arXiv preprint arXiv:1910.00927, 2019.

[42] C. Ledig, L. Theis, F. Huszár, et al., Photo-realistic single image super-resolution using a generative adversarial network, in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2017, 4681-4690.

[43] S. Liu, S. Yao, J. Li, et al., GlobalFusion: a global attentional deep learning framework for multisensor information fusion, in Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) 4, 2020, 1-27.

[44] L. Guo, L. Wang, C. Lin, et al., Wiar: a public dataset for wifi-based activity recognition, IEEE Access 7 (2019) 154935–154945.

[45] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.