

Received June 7, 2021, accepted July 4, 2021, date of publication July 8, 2021, date of current version July 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3095546

Indoor Localization Using Data Augmentation via Selective Generative Adversarial Networks

Wafa Njima¹, (Member, IEEE), Marwa Chafii¹, (Member, IEEE),
Arsenia Chorti¹, (Senior Member, IEEE), Raed M. Shubair², (Senior Member, IEEE),
and H. Vincent Poor³, (Life Fellow, IEEE)

¹ETIS UMR 8051, CY Paris Université, ENSEA, CNRS, 95000 Cergy, France.

²Department of Electrical and Computer Engineering, New York University (NYU) Abu Dhabi, Abu Dhabi, United Arab Emirates

³Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA

Corresponding author: Wafa Njima (wafa.njima@ensea.fr)

This work was supported in part by CY Initiative of Excellence (grant “Investissements d’Avenir” ANR-16-IDEX-0008) and by the project DELICATE funded by CNRS/INS2I.

ABSTRACT Several location-based services require accurate location information in indoor environments. Recently, it has been shown that deep neural network (DNN) based received signal strength indicator (RSSI) fingerprints achieve high localization performance with low online complexity. However, such methods require a very large amount of training data, in order to properly design and optimize the DNN model, which makes the data collection very costly. In this paper, we propose generative adversarial networks for RSSI data augmentation which generate fake RSSI data based on a small set of real collected labeled data. The developed model utilizes semi-supervised learning in order to predict the pseudo-labels of the generated RSSIs. A proper selection of the generated data is proposed in order to cover the entire considered indoor environment, and to reduce the data generation error by only selecting the most realistic fake RSSIs. Extensive numerical experiments show that the proposed data augmentation and selection scheme leads to a localization accuracy improvement of 21.69% for simulated data and 15.36% for experimental data.

INDEX TERMS Indoor localization, received signal strength indicator (RSSI), deep neural network (DNN), generative adversarial network (GAN), semi-supervised learning.

I. INTRODUCTION

The proliferation of smart wireless devices leveraging ubiquitous connectivity, particularly the internet of things (IoT), has driven increased interest in a wide range of services including indoor localization [1], [2]. In particular, the availability of location information has become a key factor in location-based services such as healthcare, homecare, monitoring, tracking, etc [3], [4]. To provide user location information, different wireless and emerging technologies (e.g. Bluetooth [5], WiFi [6], ultra wide-band (UWB) [7], visible light [8], magnetic field [9]) and signal parameters (e.g. time of arrival (ToA) [10], angle of arrival (AoA) [11], channel state information (CSI) [12], and received signal strength indicator (RSSI) [13]) have been investigated.

Traditional indoor localization systems are mainly based on geometric [14] (e.g. trilateration and triangulation) and fingerprinting-based methods [15]. The performance of

geometric mapping are heavily affected by multipath propagation effects which blur increasingly the relations between physical measurements and distances. Consequently, propagation modeling becomes complicated and localization accuracy is degenerated. As an alternative to geometric methods, fingerprinting-based methods have been proposed which adopt a pattern-matching process. Such methods begin by a site survey task collecting signal features at training positions in the area of interest to build a fingerprint database called also a radio map. Localization is then performed online calculating the similarity between the fingerprint at an unknown location and the fingerprint database. The estimated location corresponds to the best-fitted fingerprint applying different estimation algorithms. Classical similarity evaluation and estimation algorithms are very demanding in terms of energy and computation time, since the whole process is online and it requires to browse huge collected fingerprint databases. To overcome these issues, machine learning (ML) based indoor localization systems have been proposed [16], in particular deep learning (DL) methods [17]–[19].

The associate editor coordinating the review of this manuscript and approving it for publication was Kegen Yu¹.

Using such methods, a saved DL model based on collected fingerprint database, trained offline, is used online for accurate information location prediction which minimizes the online complexity. A variety of localization systems based on DL methods (e.g. for collected data completion [20], noise minimization [21], and location estimation [22]) have been proposed in the literature and provide good localization performance. However, in order to build an efficient DL model, these methods are data intensive, requiring very large amounts of training data, the acquisition of which is a time and energy consuming task. To address this issue, some recent approaches that leverage semi-supervised learning have been introduced, combining a small set of collected labeled data and a large set of collected unlabeled data [23]–[25]. Nevertheless, collecting unlabeled data can still be expensive, and thus, in this paper we investigate using generative models [26]–[28] for data augmentation without any extra data collection.

In the last few years, generative models have attracted significant interest in the research community due to their promising benefits in different fields. Generative models have shown a good ability to produce realistic supplement various types of data such as images, texts and sounds. In particular, generative adversarial networks (GANs) have been employed [29]–[31] to generate additional measurements, which reduces collection time and saves human effort. Such networks generate samples with improved diversity and expand the training database in order to ensure a proper design of deep neural networks (DNNs) in different fields including localization. In [32], GANs are used with semi-supervised learning where GANs use both labeled and unlabeled data to share weights with a localization classifier in order to benefit from data contained in unlabeled information when labeled data are not sufficient. The authors in [33] aim to construct an efficient radio map covering free space (e.g. open spaces and corridors) and constrained spaces. In this experiment, the data corresponding to the free space environment are measured, whereas the data for the constrained space are artificially generated. This is justified by the difficulties in performing measurement under space constraints. In the works described in [34], and [35], the initial training fingerprint database is expanded increasing the amount of training data collected at each reference point based on GANs in order to enhance its richness. For this, a collection of signal measurements have been done at different reference points and GANs have been used in order to generate additional measurements at each reference point to enhance the diversity of collected data at each reference position.

In this paper, unlike [32], we do not assume the availability of unlabeled data to enhance the training of the generative model, and we do not assume having sufficient collected data for particular regions as assumed in [33] for free space. We consider an extreme and realistic case where only a small amount of labeled data is available. We also do not use GANs to increase the diversity of the RSSI data for specific known

reference positions as proposed in [34], but we generate RSSI data for new unknown positions to cover the whole environment. In addition, we propose in this work to apply a selection process of the generated fake data that considers constraints on both the coverage and the credibility of the generated data. The proposed algorithm is validated and tested on simulated environments using realistic propagation model parameters measured from experiments. To support simulation results, we apply our proposed system on real measurements from the public UJIndoorLoc database [36].

The remainder of this paper is organized as follows: Section II formulates the localization problem and describes the system model. The proposed selective semi-supervised GANs for data augmentation is presented in Section III. The environmental setting and obtained simulation results are provided and discussed in Section IV. Experimental results based on real measurements from the UJIndoorLoc database are presented in Section V. Finally, Section VI concludes our work.

II. INDOOR LOCALIZATION SCENARIOS AND FINGERPRINTING TECHNIQUE

In this section, we briefly describe the limitations of existing indoor localization systems based on RSSI fingerprints. And, we detail different steps of the proposed system model developed to address the problem of interest.

A. SYSTEM MODEL DESCRIPTION

We consider an indoor environment covering $(L \times W)$ m², where WiFi technology is deployed using M access points (APs). As depicted in Fig. 1, the system consists of two main parts: the central unit (CU) which performs all the data processing and localization, and the mobile sensor nodes. There are two types of mobile sensor nodes: those used during the training phase in order to collect RSSI measurements for the training database construction and those requesting online to be localized. To enhance the richness of the fingerprint RSSI database, GANs are used to generate fake RSSI fingerprints. Let p_{ij} be the RSSI measured at the i -th position of the signal transmitted by the j -th AP.

B. FINGERPRINTING BASED LOCALIZATION

In this section, we present the standard fingerprinting technique when used with DNN models. Then, we detail briefly our system model steps including data collection and data training.

1) DNN-BASED FINGERPRINTING

RSSI-based positioning systems can be divided into two categories: geometric methods and fingerprints-based methods. Among them, fingerprinting is widely studied and adopted because of its localization accuracy and implementation simplicity. It consists of two main phases: an offline phase called also a training phase and an online phase called a test phase. During the offline phase, RSSI measurements are collected at each measurement location (training position)

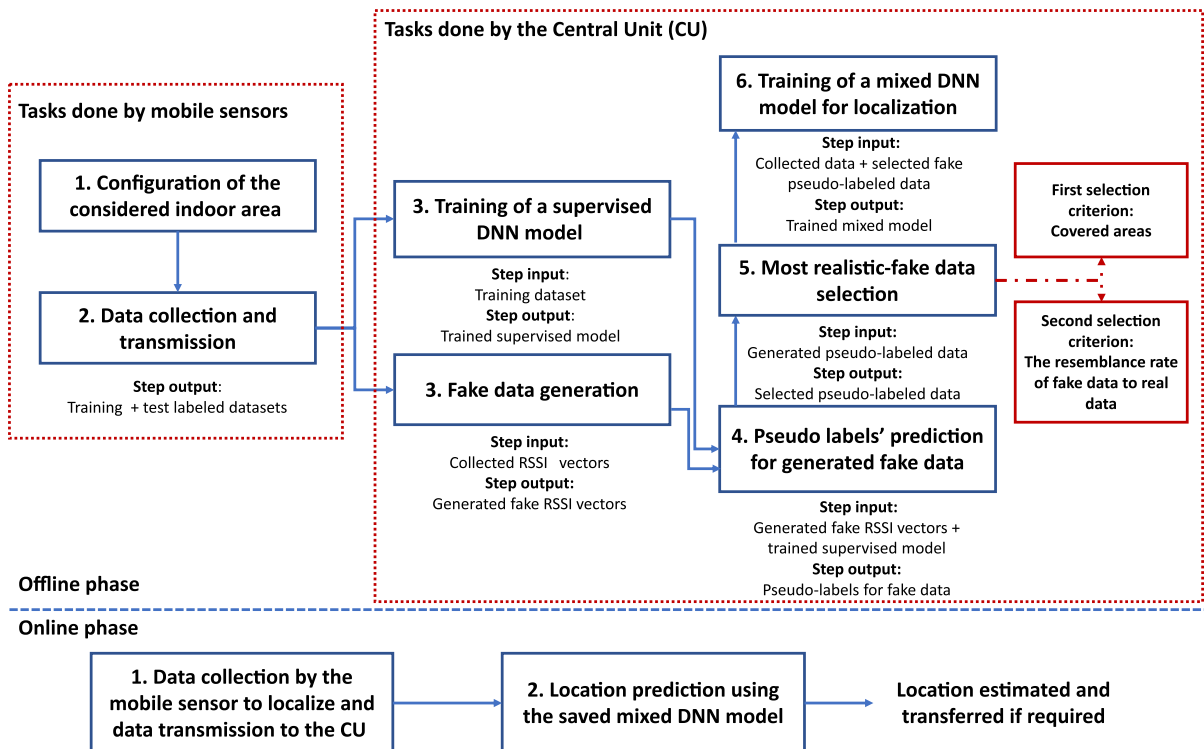


FIGURE 1. The pipeline of the proposed selective GANs based localization method during the training phase.

from different radio signal transmitters, essentially APs when considering WiFi signals. These collected RSSIs with their location coordinates are transferred to a CU and stored in a training database. In the online phase, RSSIs from different APs are measured and compared with the ones collected and stored in the training database to predict the location of the target. This prediction can be performed by a DNN model which can be trained offline and applied directly online. The use of DNN models for location prediction is recommended since most of the online complexity is shifted to the offline phase. DNN is recommended for accurate indoor localization due to the fact that it is able to learn signal fluctuations through time and environmental dynamicity because of its deeper functions that map the input to the output [37], [38]. In order to achieve good localization accuracy using an optimal trained DNN model, a large amount of expensively collected data is required during the offline phase which makes the fingerprinting method labor intensive and time/cost consuming. To solve this issue, a selective GAN indoor localization framework combined with semi-supervised learning is proposed and detailed in this paper, which performs location prediction based on collected labeled data and fake generated pseudo-labeled data. This system takes advantage of generating fake data and mixing it with labeled data in the training process in order to reduce the reliance on expensively collected data. However, generated data can be unrealistic and can create imbalanced classes when the generated data happens to be more dense for specific regions. Therefore,

a proper data selection protocol should be developed to overcome these issues.

2) DATA COLLECTION AND DATA TRAINING

In this part, we present in details the collection of labeled data, as well as the whole process of collected data pre-processing and training.

- Fingerprint database collection:** In this paper, we consider a noisy indoor environment as the area of interest in order to collect RSSI fingerprints during the offline phase including training and testing data.
- Data augmentation:** Data augmentation is used to supplement data when it is too expensive and rare. Thus, it aims to increase the size and the diversity of a dataset by generating new fake samples based on real ones.
 - Data generation:** Based on collected RSSI vectors, extra fake data are generated using a GAN in order to expand the dataset and ensure its diversity.
 - Data pseudo-labeling:** Once fake RSSI vectors are generated, an artificial pseudo-label is associated with each vector using semi-supervised learning.
- Data selection:** After data generation and pseudo-labeling, we apply a data selection method. We first generate a very large number of fake samples to cover the whole environment, and then we eliminate unnecessary and inaccurate RSSI data in order to improve the localization accuracy.

- d. **Localization based on DNNs:** To localize a target, a trained DNN model is applied to estimate its coordinates based on the corresponding collected RSSI vector. During the training process, all of the RSSI vectors (e.g, real collected vectors and selected fake generated vectors) are fitted to the model as inputs and the corresponding coordinates are considered as outputs.

III. PROPOSED SELECTIVE SEMI-SUPERVISED GAN FOR LOCATION DATA AUGMENTATION

In this section, we describe in detail different steps of the proposed localization system.

A. DATA GENERATION

1) INTUITION BEHIND GANs

Generative models aim at learning the true data distribution of a training set in order to generate realistic new samples with some variations. Thus, we try to produce samples with a distribution that is as similar as possible to the true data distribution. GANs are generative models that include two components [39]: the generator and the discriminator, as shown in Fig. 2. The generator model takes random noise as input and learns how to produce a realistic output representation similar to the real data, while the discriminator learns how to distinguish between fake and real samples. These two models are trained together until the generator is able to generate realistic examples from input noise. GANs have achieved impressive performance across a multitude of tasks (e.g. face generation, 3D object generation, and image translation from one domain to another) and many companies are using them including Google for text generation, IBM for data augmentation, Adobe for next generation photoshop and Snapchat-Tiktok for image filters.

GANs are used for data augmentation in our case, which means the generated data can be used to supplement real data to increase dataset size and diversity by using generated samples. The goal is to augment the dataset when real data are too expensive to collect.

2) GAN TRAINING

GANs consist of two different neural networks: the generator and the discriminator. To train a GAN, we alternate the training of the DNN generator model G and the DNN

discriminator model D . For each DNN model, we introduce an input vector $i_{(0)} \in \mathbb{R}^{N_0 \times 1}$ and its associated output vector $o \in \mathbb{R}^{N_{H+1} \times 1}$. Let N_h be the number of neurons for the h^{th} layer, such that $0 \leq h \leq H + 1$ where H is the number of hidden layers. $b_h \in \mathbb{R}^{N_h \times 1}$ and $W_h \in \mathbb{R}^{N_h \times N_{h-1}}$ denote the biases and the weights matrices, respectively. The output vector of the h^{th} layer can be expressed as

$$o_{(h)} = g_h(b_{(h)} + W_{(h)}i_{(h-1)}), \quad i_{(h)} = o_{(h)}, \quad (1)$$

where the input vector $i_{(h)}$ undergoes a linear transformation represented by $W_{(h)}$, a bias vector $b_{(h)}$, and then a nonlinear activation function $g_{(h)}$ is applied. During DNN training, the loss function $L(\theta) = L(W, b)$ is calculated in order to update iteratively the DNN parameters $\theta = (W, b)$.

- **Training the generator:** Let $Z \in \mathbb{R}^{M \times m_f}$ be the input dataset of the generator, such that each column $z^{(i)} \in \mathbb{R}^{M \times 1}, i = 1, \dots, m_f$ corresponds to a random noise vector whose samples are uniformly distributed over $[-1, 1]$. At the output of the DNN generator, each input noise vector $z^{(i)}$ produces a fake RSSI vector $G(z^{(i)}) \in \mathbb{R}^{M \times 1}$. The fake output vector is generated following (1), where $i_{(0)} = z^{(i)}, N_0 = M, o_{(H+1)} = G(z^{(i)})$ and $N_{H+1} = M$. Then, $G(z^{(i)})$ is passed to the DNN discriminator $D(\cdot)$ which makes a prediction of how realistic $G(z^{(i)})$ is. The generator loss is then calculated by minimizing the loss function:

$$\mathcal{L}_G = \nabla_{\theta_g} \left[-\frac{1}{m_f} \sum_{i=1}^{m_f} \log \left(D(G(z^{(i)})) \right) \right]. \quad (2)$$

Once the loss function is minimized, and the parameters $\theta_g = (W_g, b_g)$ do not change significantly after further iterations, we can save the generator DNN parameters.

- **Training the discriminator:** The discriminator learns to distinguish real from fake. At the beginning of the training, it does not know which vectors are real and which ones are fake. However, it has access to real data in order to compare them with input vectors and classify the input vectors as real or fake. Thus, the discriminator receives more and more realistic examples at each round from the generator till the examples are good enough to fill the discriminator. The discriminator acts as a binary classifier and computes a probability of an example being fake $D(G(z^{(i)}))$, following (1), giving

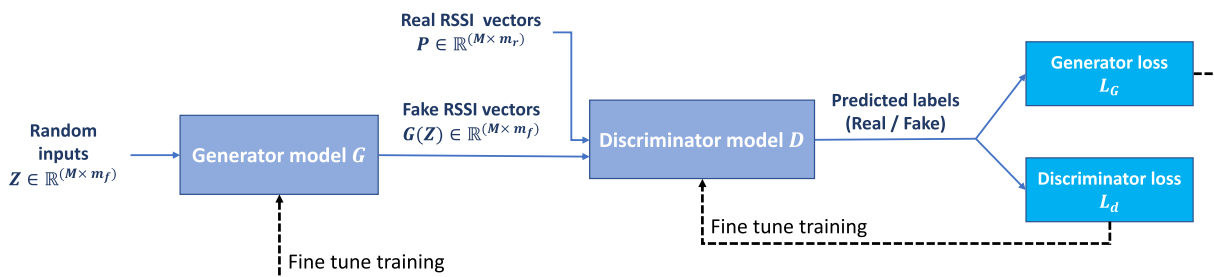


FIGURE 2. GAN network for fake RSSI data generation during training.

both fake generated samples $G(Z)$, and the real dataset $P \in \mathbb{R}^{M \times m_r}$ of collected m_r RSSIs vectors denoted as $p^{(i)}, i = 1, \dots, m_r$. This probability will be given to the generator to improve its performance as expressed in (2). The training of the discriminator is performed by minimizing the loss function derived from the binary cross-entropy (BCE) between the real and generated data [40]:

$$\mathcal{L}_d = \nabla \theta_d \left[-\frac{1}{m_r} \sum_{i=1}^{m_r} \log(D(p^{(i)})) - \frac{1}{m_f} \sum_{i=1}^{m_f} \log(1 - D(G(z^{(i)}))) \right], \quad (3)$$

where $\theta_d = (W_d, b_d)$ are the parameters of the discriminator DNN model $D(\cdot)$. $\log(D(p^{(i)}))$ refers to the log probability that the discriminator is correctly classifying the real examples while maximizing $\log(1 - D(G(z^{(i)})))$ would help the discriminator to correctly label the fake sample that comes from the generator.

3) FAKE DATA GENERATION BASED ON GANs

Given a small number of training labeled data samples, extra fake RSSI vectors are generated based on the GAN. The generator takes as inputs noise vectors that are going to be fitted to the discriminator with the real RSSI vectors. Based on the discriminator's outputs, the generator and the discriminator models are updated in order to enhance its performance generation and classification. We generate m_f fake RSSI vectors $\in \mathbb{R}^{(M \times 1)}$ based on the dataset of collected RSSI vectors, m_f is fixed experimentally in order to choose the best number of generated vectors that gives the best localization improvement.

B. SEMI-SUPERVISED FOR PSEUDO-LABELING OF GENERATED DATA

Pseudo-labeling aims to estimate the labels of an unlabeled dataset given a DNN model trained on a labeled one. In the localization context, the labels correspond to the users' location information (a room ID, a floor ID, a zone identifier, 2-D / 3-D coordinates, etc.). The 2-D coordinates are the used labels in our work. The steps of the pseudo-labeling process are summarized as follows:

- Step 1: The DNN model is trained on labeled RSSI fingerprints only, in a supervised way.
- Step 2: Based on the trained DNN model, 'pseudo-labels' are predicted for generated unlabeled RSSI vectors.
- Step 3: A mixed DNN model is trained combining labeled and selected pseudo-labeled data. To train such a model, the collected and selected RSSI vectors are used as inputs, and the corresponding associated labels (real labels and artificial pseudo-labels) are used as outputs.

During the fake data generation step, we produce unlabeled RSSI vectors. In order to be integrated into our localization

process, these fake vectors are going to be pseudo-labeled and selected. For this, collected RSSI vectors are used as inputs to a supervised DNN model that takes as outputs the corresponding labels, e.g. coordinates during training. Once trained, fake generated RSSI vectors are given to the model to predict the associated artificial pseudo-labels. Then, a general model used for localization is trained on real and selected data. The selection procedure is described in detail in subsection III-C.

C. SELECTION CRITERIA OF GENERATED FAKE DATA

Initially, the idea of this work was to generate RSSI vectors, to generate more data and cover the whole environment, and then to estimate their pseudo-labels to build a mixed and rich dataset. We have generated a large amount of fake data and tested the resulting system. However, when plotting the pseudo labels of the generated RSSIs, i.e. positions, we have noticed that (i) generated positions do not cover the whole area of the considered environment, (ii) these positions are sometimes condensed in a specific area, and (iii) generating a higher number of fake data does not necessarily lead to a better localization accuracy. Therefore, we introduce here selection criteria in order to choose only a useful subset of the generated fake data:

- **Criterion 1: Environment coverage:** Here, we consider that our environment E is uniformly divided into zones e_j each one covering $(l_j \times w_j) \text{ m}^2$, such that $\sum_j l_j \times w_j = L \times W \text{ m}^2$. After randomly generating a large amount of data (RSSI vectors and the pseudo positions), we build a new dataset by selecting for each zone e_j a number N_j of data samples proportional to its surface area $(l_j \times w_j) \text{ m}^2$, i.e. the pseudo positions fall into the desired zone e_j . Thus, redundant data are eliminated, and the whole environment E is covered and each zone e_j is equally represented in the new dataset that contains $m_s = \sum_j N_j$ selected sample. We have first tried to select randomly N_j samples per zone e_j , but we have realized that this random selection can be improved by selecting only the most realistic m_s fake data, which leads us to Criterion 2.
- **Criterion 2: Most realistic fake data:** The selection of the N_j most realistic fake data samples per zone e_j is performed by comparing the score of each generated RSSI $G(z^{(i)})$ at the output of the discriminator $D(G(z^{(i)}))$. Thus, in each zone e_j , we select the N_j positions that are most likely to be real, i.e. associated with the N_j lowest loss function values:

$$\mathcal{L}_d^{(i)} = \nabla \theta_d \left[-\frac{1}{m_r} \sum_{j=1}^{m_r} \log(D(p^{(j)})) - \log(1 - D(G(z^{(i)}))) \right], \quad (4)$$

where $p^{(j)}, j = 1, \dots, m_r$ is the j^{th} real collected RSSI vector.

IV. RESULTS BASED ON SIMULATED DATA

In this section, we present the environmental setup and different used DNN architectures and hyper-parameters. We also present the localization performance of our proposed system based on selective GANs and compare it with other methods in order to showcase its usefulness in terms of localization accuracy and data collection cost on a simulated environment.

A. PROPAGATION ENVIRONMENT AND ENVIRONMENTAL SETUP

To model the propagation environment, we consider that signals can be degraded and blocked by the obstacles:

- **Blockage:** 40% of RSSI measured data are unknown. We assign the value of -110 dBm to non-detected APs in order to have a weak signal that does not affect the calculation process. The choice of this value is based on several experiments in our simulated indoor environments.
- **Degradation** is considered combining pathloss and shadowing effects.

Let p_{ij} be the RSSI measured at the i -th position of the signal transmitted by the j -th AP. It can be expressed as

$$p_{ij} = p_t - p_{L_{ij}} + B_{\sigma} \text{ [dBm]}, \quad (5)$$

where p_t is a constant transmitted power, B_{σ} is a zero-mean Gaussian distributed random variable with standard deviation σ representing the shadowing effects, and $p_{L_{ij}}$ is the pathloss calculated as follows:

$$p_{L_{ij}} = p_{L_0} + 20 \log_{10}(f) + 10\mu \log_{10}\left(\frac{d_{ij}}{d_0}\right), \quad (6)$$

where p_{L_0} denotes the pathloss value at a reference distance d_0 , f the carrier frequency, μ the pathloss exponent whose value characterizes a specific environment and can be calculated empirically based on collected measurements and d_{ij} is the distance between the j -th AP and the i -th position.

In this part, we consider a sensor network composed of $M = 10$ APs placed randomly in an indoor environment covering $L \times W = 400 \text{ m}^2$. In order to evaluate the localization accuracy, we use a training database composed of $m_r = 1000$ RSSI vectors collected at 100 positions labeled by their coordinates. At each labeled position, we collect 10 RSSI measurements which helps to minimize temporal RSSI fluctuations caused by shadowing and fading effects. These positions are distributed in a uniform way as illustrated in Fig. 3. We divide the indoor environment covering $L \times W \text{ m}^2$ into zones/classes e_j of size $l_j \times w_j$ such that $\forall j, l_j = w_j = l$ and place a labeled position in the center of each zone in order to cover uniformly the whole area. RSSI vectors are constructed as described above. We detail in Table 1 the parameters used for the propagation model and the simulation parameters. The test database is built using the same propagation model as the one used for training data construction, but with different trajectories. We consider $m_t = 8000$ test labeled RSSI fingerprints collected at 800 test positions. In order to have a test dataset representative of

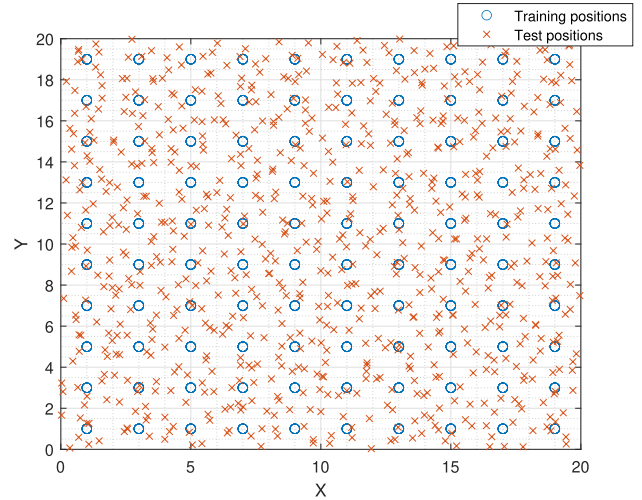


FIGURE 3. The location of training and test positions on the simulated indoor environment covering 400 m^2 .

TABLE 1. Different parameters used during the simulation process.

Parameter	Definition	Value
p_t	Transmitted power	20 dBm
f	Frequency	2.4 GHz
μ	Pathloss exponent	3.23
d_0	Reference distance	1 m
$B_{\sigma_{ij}}$	Shadowing value	$\sim \mathcal{N}(0, 4)$
L	Length of the indoor environment	20 m
W	Width of the indoor environment	20 m
l	Length of each zone	2 m
M	Number of APs	10
m_r	Number of labeled training measurements	1000
m_f	Number of generated fake positions	40000
m_t	Number of test measurements	8000

the whole environment, we place randomly each two test positions in each zone of size 1 m^2 as presented with red crosses in Fig. 3.

B. DATA GENERATION MODELS

In this section, we detail the architectures and training parameters used by different DNN models during the data generation process.

1) GAN ARCHITECTURES AND PARAMETERS

Generating useful fake data is not straightforward. For example, when generating $m_f = 1000$ fake RSSI vectors based on $m_r = 1000$ real RSSI vectors, and predicting their pseudo-positions, we have realized that the generated data do not cover the whole environment as depicted in Fig. 4. Thus, we generate a very large number of RSSI vectors in order to have pseudo-positions covering the whole area. Based on

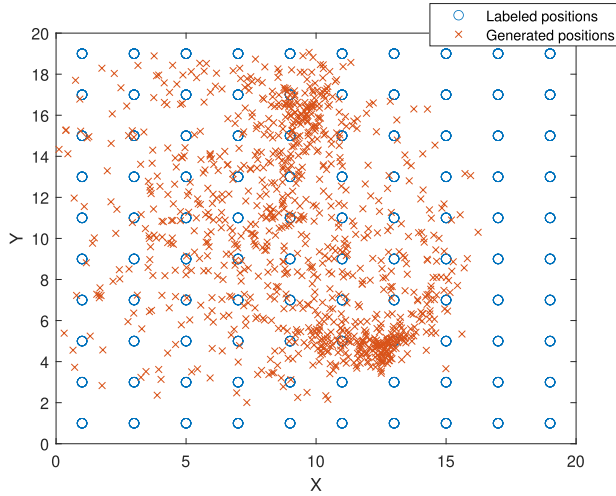


FIGURE 4. 1000 generated positions based on 100 training labeled positions.

extensive simulations, generating $m_f = 40000$ fake samples seems to cover our environment of 400 m^2 . For example, when generating $m_f = 10000$ fake data samples, we cover only 362 zones of size 1 m^2 each while with $m_f = 40000$, we cover 396 zones of size 1 m^2 as can be seen in Fig. 5a. In order to improve the diversity of the GAN, we do not generate 40000 RSSI vectors at once, but we generate 10 times 4000 fake samples based on the same real samples. In this paper, we introduce GANs based on a DNN optimized with adaptative moment estimation (ADAM) using 0.01 as the learning rate during 200 epochs. The activation function used by G is the rectified linear unit (ReLU) function used in one hidden layer having 10 neurons. The activation function of D is the ReLU function, while the last layer uses the sigmoid function. A one-hidden-layer discriminator with 10 neurons is used. These choices are based on several experiments and tests.

2) PSEUDO-LABELING ARCHITECTURES AND PARAMETERS

Once all of the data is generated, a DNN is trained based on labeled data. This model is then used to predict pseudo-labels

for generated RSSI vectors. We use ADAM as the optimization algorithm [41] for all DNN models used in this work whether for pseudo-labeling or localization. The intensive experiments have led to a DNN architecture of 2 hidden layers with 30 neurons in the first layer and 20 neurons in the second one using 200/250 epochs and a mini batch size equal to 50/100. Once generating and pseudo-labeling $m_f = 40000$, 1000 are selected, by selecting samples per covered zone of size 4 m^2 . For each covered zone, the selected samples are associated with the most realistic fake RSSIs. Note that if we skip Criterion 1, and we select directly the 1000 most realistic fake data samples, we will end up with samples not uniformly distributed over the whole environment as presented in Fig. 5b.

C. DNN MODELS USED FOR LOCALIZATION: ARCHITECTURES AND TRAINING PARAMETERS

For evaluation purposes, we compare the localization accuracy combining labeled and selected fake generated data. We called each algorithm depending on the nature and the number of used samples as follows:

- Supervised(m_r, m_p): when using a supervised method based on m_r labeled data collected at m_p different positions.
- Selective-SS-GAN(m_r, m_p, m_s) is the localization method where we combine m_r labeled samples collected at m_p different positions, and m_s pseudo-labeled selected fake generated samples for localization.

To apply localization, we need to train a DNN model on labeled collected data and pseudo-labeled data. This model takes the RSSI vectors as inputs and the outputs are the corresponding labels which are the positions' coordinates. A learning rate equal to 0.01 has been selected using 250 epochs and a mini batch size equal to 100. The DNN architectures used are summarized in Table 2 where $\text{FC}_i(\cdot)$ refers to the number of neurons in the i^{th} fully connected layer.

D. LOCALIZATION ACCURACY

In this section, we present the localization accuracy of our proposed system based on augmented dataset and we

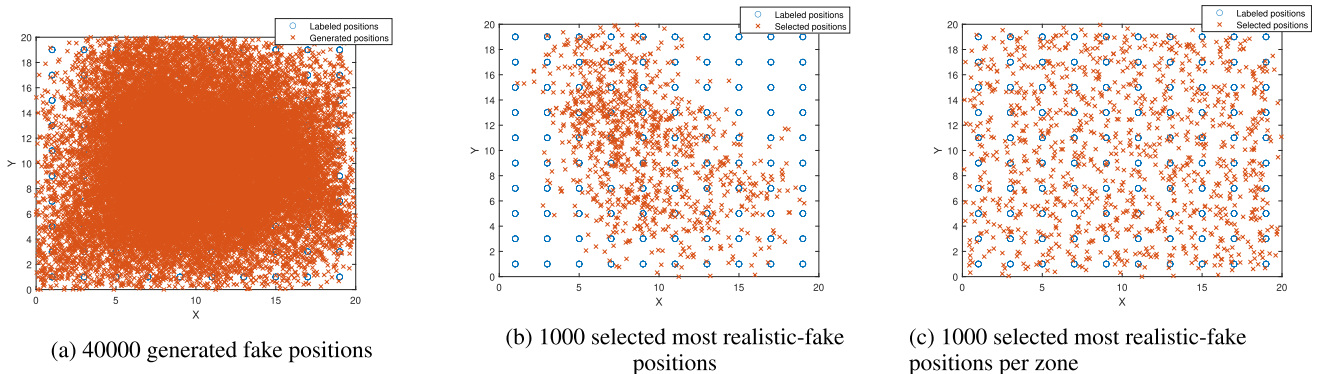


FIGURE 5. Generated positions and selected positions based on defined selection criteria.

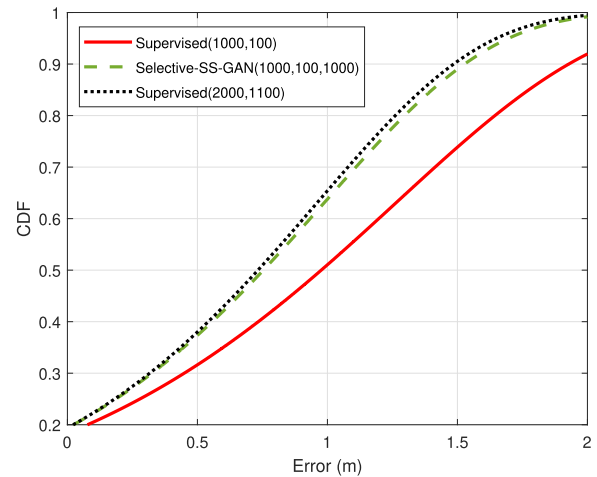
TABLE 2. Obtained localization performance considering 1000 labeled data samples.

	Localization accuracy	Accuracy increase vs supervised(1000,100)	Architecture of the DNN used for localization
Supervised(1000,100)	1.06 m	–	FC ₁ (30) and FC ₂ (20)
Selective-SS-GAN (1000,100,100)	0.87 m	19 cm 17.92%	FC ₁ (100),FC ₂ (80) and FC ₃ (20)
Selective-SS-GAN (1000,100,500)	0.84 m	22 cm 20.75%	FC ₁ (100),FC ₂ (50) and FC ₃ (10)
Selective-SS-GAN (1000,100,1000)	0.83 m	23 cm 21.96%	FC ₁ (80),FC ₂ (60) and FC ₃ (20)
Selective-SS-GAN (1000,100,2000)	0.85 m	21 cm 19.81%	FC ₁ (100),FC ₂ (80) and FC ₃ (20)
Selective-SS-GAN (1000,100,3000)	0.86 m	20 cm 18.86%	FC ₁ (80),FC ₂ (60) and FC ₃ (20)
Selective-SS-GAN (1000,100,4000)	0.86 m	20 cm 18.86%	FC ₁ (80),FC ₂ (60) and FC ₃ (20)
Supervised(2000,1100)	0.81 m	25 cm 23.58%	FC ₁ (30) and FC ₂ (20)

compare it with the scheme where only labeled collected data samples are used for localization. Results corresponding to these data are shown in Table 2.

In Table 2, we present the localization accuracy (i.e., localization mean error) for Selective-SS-GAN(m_r , m_p , m_s) trained over $m_r = 1000$ labeled data samples and different numbers of selected fake data samples vs. supervised learning model trained only on $m_r = 1000$ labeled samples. We select different numbers of fake data samples [100 – 4000] from the set of $m_f = 40000$ generated samples while satisfying Criterion 1 and Criterion 2. We notice that for all augmented datasets, the localization accuracy is improved compared to a dataset limited only to real data. The localization improvement varies between 17.92% and 21.96%. The best localization accuracy is obtained with $m_s = 1000$ selected fake data samples, where we achieve 21.96% localization accuracy increase vs. the conventional supervised algorithm without any additional cost in collecting additional data. This improvement is explained by the consideration that the DNN has been trained over a larger dataset which covers new positions that are not observed in the limited dataset based on collected data. Moreover, we can see that generating more fake data does not necessarily improve the localization accuracy, which can be explained by the GAN generation error due to generating less realistic RSSI vectors. Note that starting from 2000 generated samples, the performance saturates and no improvement can be achieved by generating additional fake data. This can be explained by the fact that based on 1000 labeled vectors collected at 100 training positions, we cannot provide a higher measurement diversity to the GAN.

Fig. 6 illustrates the cumulative distribution functions (CDFs) of three algorithms combining data subsets of different types (labeled and generated). We can easily notice that the supervised indoor localization system based on 1000 labeled samples collected at 100 known positions corresponds to the worst localization accuracy. For fair comparison, we use the same dataset of labeled positions to which we add (i) 1000 labeled measurements collected at 1000 different labeled positions placed randomly in the considered area i.e. Supervised(2000,1100), and (ii) based on these data, we generate and select 1000 fake positions i.e. Selective-SS-GAN(1000,100,1000). We notice that the localization

**FIGURE 6.** Localization performance comparison.

performance is close with only 2 cm of difference, which means that we can achieve almost the same performance by only collecting half of the labeled data and artificially generating the other half. Thus, the proposed data generation and selection process provides an improvement of localization accuracy without additional collection cost.

V. RESULTS BASED ON EXPERIMENTAL DATA

In order to support the simulation results, we validate and test our system on real RSSI measurements from the public UIndoorLoc database. This database covers the area of almost 110000 m² including three buildings with four or five floors. For implementation simplicity and since we work on one-floor area, we consider collected data corresponding to the second floor from building 1 only. There are training measurements and validation measurements collected four months later than training ones. For the considered floor, we have 1395 training fingerprints taken at 80 training positions and 40 validation positions taken at different validation positions. For fair comparison and since we consider only 1000 training fingerprints as system input data during simulations, we choose randomly 1000 fingerprints from the training set and we add the 395 rest training positions

TABLE 3. Obtained localization performance considering 1000 real fingerprints corresponding to Building1-Floor2 from UJIndoorLoc database.

	Localization mean error	Localization min-max errors	Localization mean error decrease vs supervised(1000)	Architecture of the DNN used for localization
Supervised(1000)	4.1 m	3.92 m - 4.53 m	–	FC ₁ (200), FC ₂ (100) and FC ₃ (50)
Selective-SS-GAN (1000,700)	3.67 m	3.58 m – 3.72 m	43 cm 10.48 %	FC ₁ (300),FC ₂ (150) and FC ₃ (30)
Selective-SS-GAN (1000,1000)	3.47 m	3.38 m – 3.57 m	63 cm 15.36 %	FC ₁ (300),FC ₂ (150) and FC ₃ (50)
Selective-SS-GAN (1000,1500)	3.77 m	3.74 m – 3.79 m	33 cm 8.04 %	FC ₁ (300),FC ₂ (150) and FC ₃ (30)
Selective-SS-GAN (1000,2000)	3.83 m	3.76 m – 3.94 m	27 cm 6.58 %	FC ₁ (300),FC ₂ (150) and FC ₃ (30)

to the 40 validation positions to be used for test. During UJIndoorLoc construction, 520 APs are used. However, only 18 APs are detected at each position and 91% of collected RSSI are unknown. Consequently, for implementation simplicity, we consider only APs detected at least once during data collection which are equal to 190.

After data re-organisation and reduction, we are able to apply our algorithm. At first, we generate $m_f = 28000$ fake RSSI vectors with GANs based on DNN, optimized using 0.01 as learning rate during 200 epochs. A one-hidden layer discriminator with 200 neurons and a one-hidden layer generator with 200 neurons are considered. For the optimization algorithm and the activation functions, we use the ones used during simulations. Secondly, a DNN is trained on labeled measurements during 200 epochs with 50 as a batch size for pseudo-labels estimation. The best achievable DNN is composed by three hidden layers containing 200 neurons, 100 neurons and 50 neurons, respectively. For data selection, we consider that we split the zone of interest into 625 classes with $l = 6.4959$ m and $w = 5.9760$ m. The data selection process is followed by location estimation applying a DNN trained offline. A learning rate equal to 0.01 has been selected using 150/250 epochs and a mini batch size equal to 50/100. The DNN architectures used for localization are presented in Table 3.

We evaluate the localization accuracy of different algorithms using both real and selected fake data i.e. Selective-SS-GAN (m_f, m_s). Table 3 presents the localization performance (e.g., mean localization error, min localization error and max localization error) and the localization improvement compared of to the localization system using only 1000 labeled data i.e. Supervised (1000). We select different numbers of fake positions [700 - 2000] from the set of $m_f = 28000$ generated positions. We notice that the localization improvement varies between 6.58% and 15.36%. The best localization accuracy is obtained with $m_s = 1000$ selected generated positions, where we achieve 15.36% localization accuracy increase vs. the conventional supervised algorithm. Starting from 1500 selected generated positions, we notice that the performance is saturated and we cannot provide better accuracy based on training positions. Thus, experimental results strengthen results based on simulated data. Even when working in a realistic environment with high dynamic and heterogeneous devices, our proposed system achieves good localization accuracy.

VI. CONCLUSION

In this paper, we have seen that GANs can be used to produce synthetic data, outstandingly realistic, to complement a real dataset in order to enhance the training of a DNN used for localization. Thus, this technique is very useful in situations where data collection is expensive and time consuming such as indoor localization. In particular, we have presented a selective semi-supervised GAN system for indoor localization, where we have generated fake data based on real labeled collected data in order to boost localization performance. Our proposed solutions have been validated and tested based on several simulations which show that the combination of collected data and selected generated data is beneficial in terms of localization performance and data collection cost. The selection-generation process improves the localization accuracy by 21.69% compared to the standard supervised method based on the same subset of labeled data when considering 1000 labeled samples collected at 100 different positions. Our method works also for real data, from the public UJIndoorLoc database, leading to a localization accuracy improvement of 15.36%. The promising results of this paper motivate further extension of the developed model by exploring new methods for data augmentation.

REFERENCES

- [1] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 3rd Quart., 2019.
- [2] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs, and E. Aboutanios, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1327–1346, 2nd Quart., 2016.
- [3] J. A. Del Peral-Rosado et al., "Whitepaper on new localization methods for 5G wireless systems and the Internet-of-Things," 2018.
- [4] A. Bourdoux, A. N. Barreto, B. van Liempd, C. de Lima, D. Dardari, D. Belot, E.-S. Lohan, G. Seco-Granados, H. Srieddeen, H. Wymeersch, J. Suutala, J. Saloranta, M. Guillaud, M. Isomursu, M. Valkama, M. R. K. Aziz, R. Berkvens, T. Sanguanpuak, T. Svensson, and Y. Miao, "6G white paper on localization and sensing," 2020, *arXiv:2006.01779*. [Online]. Available: <http://arxiv.org/abs/2006.01779>
- [5] S. Liu, Y. Jiang, and A. Striegel, "Face-to-face proximity estimation using Bluetooth on smartphones," *IEEE Trans. Mobile Comput.*, vol. 13, no. 4, pp. 811–823, Apr. 2013.
- [6] S. He and S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 466–490, Jan. 2015.
- [7] B. Kempke, P. Pannuto, and P. Dutta, "SurePoint: Exploiting ultra wide-band flooding and diversity to provide robust, scalable, high-fidelity indoor localization: Demo abstract," in *Proc. 14th ACM Conf. Embedded Netw. Sensor Syst. CD-ROM*, Nov. 2016, pp. 137–149.

- [8] A. B. M. M. Rahman, T. Li, and Y. Wang, "Recent advances in indoor localization via visible lights: A survey," *Sensors*, vol. 20, no. 5, p. 1382, Mar. 2020.
- [9] W. Shao, H. Luo, F. Zhao, Y. Ma, Z. Zhao, and A. Crivello, "Indoor positioning based on fingerprint-image and deep learning," *IEEE Access*, vol. 6, pp. 74699–74712, 2018.
- [10] N. Tadayon, M. T. Rahman, S. Han, S. Valaee, and W. Yu, "Decimeter ranging with channel state information," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3453–3468, Jul. 2019.
- [11] L. Chen, I. Ahriz, and D. Le Ruyet, "AoA-aware probabilistic indoor location fingerprinting using channel state information," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 10868–10883, Nov. 2020.
- [12] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–32, 2013.
- [13] P. Kumar, L. Reddy, and S. Varma, "Distance measurement and error estimation scheme for RSSI based localization in wireless sensor networks," in *Proc. 5th Int. Conf. Wireless Commun. Sensor Netw. (WCSN)*, Dec. 2009, pp. 1–4.
- [14] Z. Yang and Y. Liu, "Quality of trilateration: Confidence-based iterative localization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 5, pp. 631–640, May 2010.
- [15] S. Yiu, M. Dashti, H. Claussen, and F. Perez-Cruz, "Wireless RSSI fingerprinting localization," *Signal Process.*, vol. 131, pp. 235–244, Feb. 2017.
- [16] M. I. AlHajri, N. T. Ali, and R. M. Shubair, "Indoor localization for IoT using adaptive feature selection: A cascaded machine learning approach," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2306–2310, Nov. 2019.
- [17] W. Njima, I. Ahriz, R. Zayani, M. Terre, and R. Bouallegue, "Deep CNN for indoor localization in IoT-sensor systems," *Sensors*, vol. 19, no. 14, p. 3127, Jul. 2019.
- [18] W. Njima, M. Chafii, A. Nimr, and G. Fettweis, "Deep learning based data recovery for localization," *IEEE Access*, vol. 8, pp. 175741–175752, 2020.
- [19] Z. Chen, M. I. AlHajri, M. Wu, N. T. Ali, and R. M. Shubair, "A novel real-time deep learning approach for indoor localization based on RF environment identification," *IEEE Sensors Lett.*, vol. 4, no. 6, pp. 1–4, Jun. 2020.
- [20] R. Wang, Z. Li, H. Luo, F. Zhao, W. Shao, and Q. Wang, "A robust Wi-Fi fingerprint positioning algorithm using stacked denoising autoencoder and multi-layer perceptron," *Remote Sens.*, vol. 11, no. 11, p. 1293, May 2019.
- [21] J. Fan and T. Chow, "Deep learning based matrix completion," *Neurocomputing*, vol. 266, pp. 540–549, Nov. 2017.
- [22] R. Ayyalasomayajula, A. Arun, C. Wu, S. Sharma, A. R. Sethi, D. Vasishth, and D. Bharadia, "Deep learning based wireless localization for indoor navigation," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, Apr. 2020, pp. 1–14.
- [23] W. Li, C. Zhang, and Y. Tanaka, "Pseudo label-driven federated learning-based decentralized indoor localization via mobile crowdsourcing," *IEEE Sensors J.*, vol. 20, no. 19, pp. 11556–11565, Oct. 2020.
- [24] Y. Xia, L. Ma, Z. Zhang, and Y. Wang, "Semi-supervised positioning algorithm in indoor WLAN environment," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.
- [25] M. Zhou, Y. Tang, W. Nie, L. Xie, and X. Yang, "GrassMA: Graph-based semi-supervised manifold alignment for indoor WLAN localization," *IEEE Sensors J.*, vol. 17, no. 21, pp. 7086–7095, Nov. 2017.
- [26] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," 2017, *arXiv:1711.04340*. [Online]. Available: <http://arxiv.org/abs/1711.04340>
- [27] C. Bowles, R. Gunn, A. Hammers, and D. Rueckert, "GANsfer learning: Combining labelled and unlabelled data for GAN based data augmentation," 2018, *arXiv:1811.10669*. [Online]. Available: <http://arxiv.org/abs/1811.10669>
- [28] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," 2018, *arXiv:1809.11096*. [Online]. Available: <http://arxiv.org/abs/1809.11096>
- [29] A. Odena, "Semi-supervised learning with generative adversarial networks," 2016, *arXiv:1606.01583*. [Online]. Available: <http://arxiv.org/abs/1606.01583>
- [30] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, D. A. Dickie, M. V. Hernández, J. Wardlaw, and D. Rueckert, "GAN augmentation: Augmenting training data using generative adversarial networks," 2018, *arXiv:1810.10863*. [Online]. Available: <http://arxiv.org/abs/1810.10863>
- [31] A. Belmonte-Hernández, G. Hernandez-Penaloza, D. M. Gutiérrez, and F. Alvarez, "Recurrent model for wireless indoor tracking and positioning recovering using generative networks," *IEEE Sensors J.*, vol. 20, no. 6, pp. 3356–3365, Mar. 2020.
- [32] K. M. Chen and R. Y. Chang, "Semi-supervised learning with GANs for device-free fingerprinting indoor localization," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.
- [33] H. Zou, C.-L. Chen, M. Li, J. Yang, Y. Zhou, L. Xie, and C. J. Spanos, "Adversarial learning-enabled automatic WiFi indoor radio map construction and adaptation with mobile robot," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 6946–6954, Aug. 2020.
- [34] Q. Li, H. Qu, Z. Liu, N. Zhou, W. Sun, S. Sigg, and J. Li, "AF-DCGAN: Amplitude feature deep convolutional GAN for fingerprint construction in indoor localization systems," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 3, pp. 468–480, Jun. 2021.
- [35] Y. Lei, D. Li, H. Zhang, and X. Li, "Wavelet feature outdoor fingerprint localization based on ResNet and deep convolution GAN," *Symmetry*, vol. 12, no. 9, p. 1565, Sep. 2020.
- [36] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2014, pp. 261–270.
- [37] K. S. Kim, R. Wang, Z. Zhong, Z. Tan, H. Song, J. Cha, and S. Lee, "Large-scale location-aware services in access: Hierarchical building/floor classification and location estimation using Wi-Fi fingerprinting based on deep neural networks," *Fiber Integr. Opt.*, vol. 37, no. 5, pp. 277–289, Sep. 2018.
- [38] K. S. Kim, S. Lee, and K. Huang, "A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting," *Big Data Anal.*, vol. 3, no. 1, pp. 1–17, Dec. 2018.
- [39] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*. [Online]. Available: <http://arxiv.org/abs/1406.2661>
- [40] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for GANs do actually converge?" in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 3481–3490.
- [41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>



Wafa Njima (Member, IEEE) was born in Tunisia, in 1991. She received the degree in engineering from the Institut National des Sciences Appliquées et de Technologies, Tunis, in 2015, and the Ph.D. degree in radio-communications from the Conservatoire National des Arts et Métiers, Paris, in 2019, in collaboration with the Ecole Supérieure des Communications de Tunis, under the supervision of Prof. Michel Terré and Prof. Ridha Bouallegue. She is currently a Postdoctoral Researcher with CNRS, ETIS UMR 8051, ENSEA, CY Cergy Paris University, France. Her publications span several research areas, including signal processing, wireless communications, sparse data, indoor localization, the IoT, and machine learning for communications.



MARWA CHAFII (Member, IEEE) received the master's degree in advanced wireless communication systems (SAR) and the Ph.D. degree in electrical engineering from CentraleSupélec, France, in 2013 and 2016, respectively. From 2014 to 2016, she has been a Visiting Researcher with the Poznan University of Technology, Poland; the University of York, U.K.; Yokohama National University, Japan; and the University of Oxford, U.K. She joined the Vodafone Chair Mobile Communication Systems, Technical University of Dresden, Germany, in February 2018, as a Research Group Leader. Since September 2018, she has been an Associate Professor with ENSEA, France, where she holds the Chair of Excellence on artificial intelligence from CY Initiative. Her research interests include advanced waveform design, machine learning for wireless communications, and indoor localization. She received the prize of the Best Ph.D. in France in the fields of signal, image and vision, and she has been nominated in the top ten Rising Stars in Computer Networking and Communications by N2Women, in 2020. She is currently the Vice-Chair of the IEEE ComSoc ETI on Machine Learning for Communications, leading the Education Working Group, ETI on Integrated Sensing and Communications, a Research Lead with Women in AI, and managing the Gender Committee of the AI4EU Community. Since 2019, she has been serving as an Associate Editor for IEEE COMMUNICATIONS LETTERS, where she received the Best Editor Award, in 2020.



ARSENIA (ERSI) CHORTI (Senior Member, IEEE) is currently a Professor with the Ecole Nationale Supérieure de l'Electronique et des ses Applications (ENSEA); the Head of the Information, Communications and Imaging (ICI) Group, ETIS Lab UMR 8051; and a Visiting Research Fellow with Princeton University and the University of Essex. Her research interests include wireless communications and wireless system security for 5G and 6G, with a particular focus on physical layer security, context aware security, multi-factor authentication protocols, 5G/6G and the IoT, anomaly detection, machine learning for communications, non-orthogonal multiple access (NOMA), and faster than Nyquist signaling. She is a member of IEEE INGR on Security and the IEEE P1951.1 Standardization Workgroup (Smart Cities). She has been on a semi-sabbatical leave (CNRS delegation), since March 2021.



RAED M. SHUBAIR (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Waterloo, Canada, in 1993. From 1993 to 2017, he was a Full Professor of electrical engineering with Khalifa University, United Arab Emirates. He is currently a Full Professor of electrical engineering affiliated to New York University (NYU) Abu Dhabi. His current and past academic and research appointments also include the Massachusetts Institute of Technology (MIT), Harvard University, and the University of Waterloo. His publication span several research areas, including 6G and terahertz communications, modern antennas and applied electromagnetics, signal and array processing, machine learning, the IoT and sensor localization, medical sensing, and nano-biomedicine. He is a fellow of the MIT Electromagnetics Academy, and a Founding Member of MIT Scholars of the Emirates and five IEEE society chapters, United Arab Emirates. He received the University of Waterloo Distinguished Doctorate Dissertation Award for his Ph.D. degree. He was a recipient of several international awards, including the Distinguished Service Award from ACES Society, USA, and the MIT Electromagnetics Academy, USA. He is also a standing member of the editorial boards of several international journals and serves regularly on the steering, organizing, and technical committees of IEEE flagship conferences in antennas, communications, and signal processing, including IEEE AP-S/URSI, EuCAP, IEEE GlobSIP, IEEE WCNC, and IEEE ICASSP. He is also a Board Member of the European School of Antennas and the Regional Director for the IEEE Signal Processing Society in IEEE Region 8 Middle East, and served as the Founding Chair of the IEEE Antennas and Propagation Society Educational Initiatives Program. He is also the Founder and the Chair of IEEE at NYUAD. He is also an Officer of IEEE ComSoc Emerging Technical Initiative on Machine Learning for Communications. He is also the Founding Director of the IEEE UAE Distinguished Seminar Series Program for which he was selected to receive, along with Mohamed AlHajri of MIT, the 2020 IEEE UAE Award of the Year. He is also an Editor of the IEEE JOURNAL OF ELECTROMAGNETICS, RF AND MICROWAVES IN MEDICINE AND BIOLOGY, and the IEEE OPEN JOURNAL OF ANTENNAS AND PROPAGATION.



H. VINCENT POOR (Life Fellow, IEEE) received the Ph.D. degree in EECS from Princeton University, in 1977. From 1977 to 1990, he was with the Faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been with the Faculty at Princeton University, where he is currently the Michael Henry Strater University Professor. From 2006 to 2016, he served as the Dean of the Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other universities, including most recently at Berkeley and Cambridge. His research interests include information theory, machine learning, and network science, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the forthcoming book on machine learning and wireless communications (Cambridge University Press, 2021). He is a member of the National Academy of Engineering and the National Academy of Sciences, and a Foreign Member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal and the D.Eng. Honoris Causa from the University of Waterloo awarded, in 2019.

...