# Open-Set Occluded Person Identification With mmWave Radar

Tao Wang, *Student Member, IEEE*, Yang Zhao, *Senior Member, IEEE*, Ming-Ching Chang, *Senior Member, IEEE*, and Jie Liu, *Fellow, IEEE*

***Abstract*—Radio frequency sensors can penetrate non-metal objects and provide complementary information to vision sensors for person identification (PID) purposes. However, there is a lack of research on millimeter wave (mmWave) radar for PID under occlusions, particularly in addressing the open-set recognition problem. Thus, we propose an open-set occluded PID (OSO-PID) framework that can deal with various obstacle and occlusion scenarios with open-set recognition capability. We first introduce a new dataset, mmWave-ocPID, comprising mmWave radar measurements and RGB-depth images, collected from 23 human subjects. We next design a novel neural network, mm-PIDNet, for occluded person identification using mmWave radar measurements. mm-PIDNet incorporates a transformer encoder, a bidirectional long short-term memory module, and a novel supervised contrastive learning module to improve PID performance. For open-set recognition, we enhance the mmWave radar-based PID method by integrating supervised contrastive learning with the Weibull models, which can identify out-of-distribution samples. We perform extensive indoor experiments with a variety of obstacles and occlusion scenarios. Our experimental results show that mm-PIDNet achieves an F1-score of 0.93 on average, outperforming state-of-the-art methods by up to 13.41% for occluded cases. For open-set PID, the OSO-PID framework achieves an F1-score above 0.8 when the openness is less than 14.36%.**

***Index Terms*—Wireless sensing, mmWave radar, person identification, open-set recognition, contrastive learning.**

## I. Introduction

With the widespread adoption of mobile computing techniques, wireless sensors have been used for various human sensing tasks, such as occupancy detection, gesture recognition, vital sign monitoring, etc. [1], [2], [3], [4]. Among these sensing tasks, person identification (PID) aims to distinguish people based on unique characteristics including facial appearance, fingerprints, and gaits. It finds applications in surveillance, smart facilities, health care, etc. Conventional vision sensor-based PID approaches assume visibility of the whole body or identifiable parts [5], [6]. However, in real-world scenarios such as home and office environments, occlusions are common, leading to poor performance or complete failure of vision sensor-based PID methods.

Radio frequency (RF) sensors provide complementary information to vision sensors and non-line-of-sight (NLOS) capability for PID. Compared with other RF sensors, the millimeter-wave (mmWave) radar sensor has high portability and sensing capability due to its short wavelength and large bandwidth [7]. Furthermore, mmWave radar excels in PID under low-light conditions, adverse weather, and penetrating non-metallic objects [8], [9], [10], [11]. Thus, recent research studies use mmWave radar to extract gait, a distinctive walking pattern, to achieve satisfactory PID performance at a distance without person cooperation [3], [12], [13]. However, none of the existing mmWave radar-based works deal with PID under heavy occlusion scenarios due to the following changes.

First, in occlusion scenarios, obstacles and moving human individuals produce more complicated, time-varying multipath propagation of radar signals [14]. In addition to direct reflections from human subjects, signals with indirect reflection paths are also captured by the radar [15]. Thus, the mmWave radar signals have more complicated multipath effects in occluded person identification scenarios. Second, mmWave signals can be greatly weakened when passing through an obstacle [16]. This reduces the number of mmWave signals captured by the radar, leading to insufficient information for accurate person identification. Third, data-driven methods require extensive data for training. However, due to the laborious and time-consuming data collection process, there is no dataset publicly available for mmWave radar-based person identification under occluded conditions.

In addition, most existing radar-based PID systems operate under a closed-set assumption, where the testing data is assumed to be independently and identically distributed (*i.i.d.*) as the training data [17]. However, real-world scenarios often involve open environments, where unforeseen classes, i.e., individuals, emerge unexpectedly, significantly reducing the effectiveness of existing methods [18]. For example, for a closed-set PID system, there is a risk of incorrectly identifying a home intruder as a family member, since only data from family members are collected for training the PID system. To address this issue, we propose to formulate the PID problem as an open-set

Received 29 January 2024; revised 25 December 2024; accepted 8 January 2025. Date of publication 15 January 2025; date of current version 7 May 2025. This work was supported in part by the National Key R & D Program of China under Grant 2022YFF0503900 and in part by the National Natural Science Foundation of China under Grant 62350710797. Recommended for acceptance by L. Wang. *(Corresponding author: Yang Zhao.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Harbin Institute of Technology Ethics Committee, and our IRB under Application No. HIT-2023062.

Tao Wang, Yang Zhao, and Jie Liu are with the International Research Institute for Artificial Intelligence, Harbin Institute of Technology (Shenzhen), Shenzhen 518071, China (e-mail: zhao.yang@ieee.org).

Ming-Ching Chang is with the Department of Computer Science, College of Engineering and Applied Sciences University at Albany, State University of New York, Albany, NY 12222 USA.

Digital Object Identifier 10.1109/TMC.2025.3529735

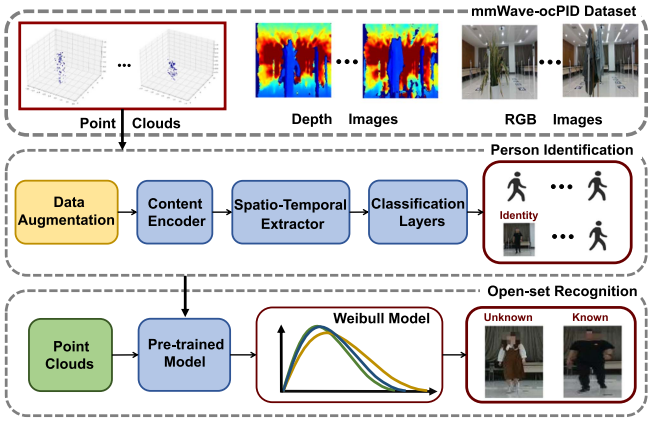



Fig. 1. An overview of the OSO-PID framework, in which a new mmWave radar PID dataset is built with various occlusion scenarios included to enable the occluded person sensing and open-set recognition capabilities.

recognition (OSR) task [17], [19], [20]. In OSR, the multi-class classifier needs to simultaneously classify test samples of the *known* classes, and recognize testing samples of the *unknown* classes. This allows the model to identify individuals based on an established individual set, while also effectively rejecting unknown intruders. While OSR methods have been proposed for vision-based tasks [21], [22], [23], there exists a research gap in OSR for mmWave radar-based PID, especially at various occlusion scenarios.

In this study, we take advantage of the penetration capability of the mmWave radar to study the open-set person recognition problem at occluded conditions. To our knowledge, this work is the first mmWave radar-based PID framework with open-set recognition capability designed for occlusion conditions. Fig. 1 overviews the OSO-PID framework, and major components are explained next.

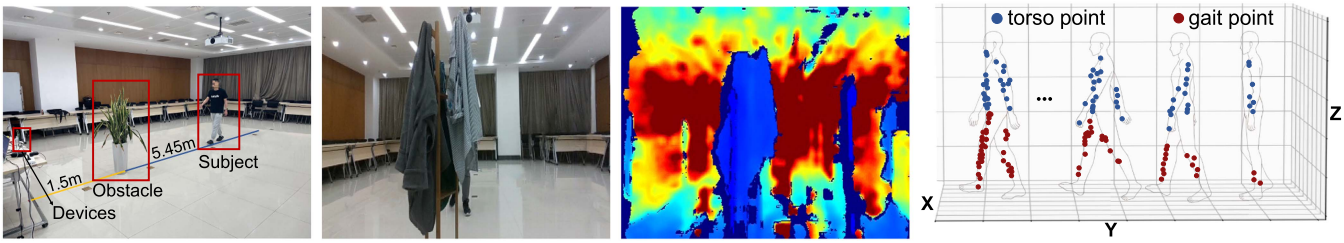
*The mmWave-ocPID dataset:* To address the data gap, we build a new dataset, mmWave-ocPID, containing mmWave radar measurements from 23 human subjects in various occlusion scenarios. We perform extensive experiments emphasizing the evaluation of mmWave radar-based PID under multiple occlusion scenarios, as shown in Fig. 2. The dataset comprises data collected from commercial off-the-shelf (COTS) sensors: (1) measurements collected using a mmWave radar, and (2) synchronized RGB and depth images collected using an RGB-D camera. The dataset is collected at various occlusion conditions, with different objects, e.g., poster boards, clothes racks and plants between human subjects and sensors. The dataset contains overall 300,000 radar frames, together with more than 600,000 RGB and depth images, providing a comprehensive resource for investigating mmWave radar-based PID under occlusion. Note that the images are collected for experiment recording and visualization purposes, they are not used in mmWave radar-based PID training or inference.

*The PID neural network model with data augmentation:* To enrich the information in radar point clouds, we propose a new data augmentation strategy that leverages the spatial-temporal relationships within radar point clouds to generate novel signatures, enhancing the performance of person identification. Then, we propose a novel neural network for person identification in occluded scenarios, called mm-PIDNet. The network, coupled with a new supervised contrastive learning module [24], takes a sequence of augmented radar point clouds as input to extract high-dimensional features and spatial-temporal relationships to achieve accurate and robust person identification under occluded scenarios. mm-PIDNet combines a content encoder, a transformer encoder [25] and a bidirectional long short-term memory (Bi-LSTM) [26] module to extract features for the identification of individuals. The supervised contrastive learning module implements an offline buffer specifically designed for contrastive samples and uses the momentum update method [27] to maintain sample consistency. It brings samples with the same class label closer in the feature space, mitigating the overfitting issue of the PID model and generating a compact feature space that benefits OSR learning.

*The open-set recognition method:* We propose a new method that combines Weibull models with mm-PIDNet for open-set occluded person identification. The method not only classifies testing samples from known classes but also recognizes testing samples from unknown classes. Our method training involves two key steps: (1) closed-set training, and (2) fitting Weibull models. During the closed-set PID network training, we incorporate supervised contrastive learning to establish a compact feature space. According to [23], the Weibull distribution function is upper bounded, which is suitable to limit the feature support region of each known class. Hence, we utilize Weibull models, fitted using distances between activation vectors [21] and class centers derived by averaging activation vectors from correctly classified training samples. Using the fitted Weibull model, the probability of each testing sample belonging to each class can be determined by measuring the distance between the feature from the pre-trained mm-PIDNet and the class center. The fitted Weibull model further refines the feature space constructed by mm-PIDNet by constraining the distance of features from the class center, and it generates probabilities for identifying known classes as well as recognizing unknown classes.

To summarize, the contributions of this paper are as follows:

- We use a COTS mmWave radar sensor to build the mmWave-ocPID dataset, featuring a diverse array of occlusion scenarios. We make the dataset publicly available in IEEE Dataport (http://ieee-dataport.org/12089), which includes over 300, 000 radar frames and over 600, 000 RGB and depth images.
- We propose a novel neural network combining transformer and Bi-LSTM architectures for PID in occluded scenarios. Enhanced with a data augmentation strategy and a supervised contrastive learning approach, these methods effectively reduce model overfitting and improve PID performance under occlusions.
- We present a novel OSR framework for mmWave radar point cloud-based PID. Supervised contrastive learning enables the construction of a compact feature space, and Weibull models enable accurate identifications of samples from both known and unknown classes.
- Experimental results show an average PID F1-score reaching 0.93 on our dataset. Compared with state-of-the-art




(a) Experiment layout (b) RGB image (c) Depth image (d) Point cloud sequence

Fig. 2. PID experimental layout and data collected in experiments. (a) Experimental layout with human subjects, obstacles, and sensing devices. (b-d) RGB image, depth image, and mmWave radar point cloud of a walking individual, respectively.

methods, our method shows an F1-score improvement of up to 13.41% in distinct occlusion cases. Additionally, our open-set method achieves an F1-score over 80% with an openness score below 14.36%.

The rest of the paper is organized as follows: Section II presents related work. Section III describes an overview of the theoretical analysis and the problem statement. Section IV illustrates our OSO-PID framework design in detail. Section V describes experimental setups and our mmWave-ocPID dataset. Section VI reports evaluation results. Section VII concludes our work.

## II. Relate Work

### A. Mmwave Radar-Based Person Identification

Person identification and re-identification are fundamental tasks, which have achieved significant success in the computer vision field [28], [29]. For example, [28] proposes a person re-identification model to learn resolution-adaptive representations for recognizing individuals using images collected from various resolutions. [29] considers adversarial attacks in person re-identification and develops a robust model using generative metric learning to enhance identification performance. Although these computer vision-based methods have achieved satisfactory results, their performance could be degraded under heavy or full occlusion, as it affects the ability of the camera to capture information about the whole body or identifiable parts of the target. In this paper, we address this issue by using a radio frequency (RF) sensor, mmWave radar, which can penetrate common obstacles for person identification.

Gait analysis-based PID methods using mmWave radar features can be categorized into two types: Dopper and point cloud methods. Dopper-based techniques, such as those in [30], [31], [32], capture micro-Dopper signatures from human movement. These signatures are transformed into visual representations, which can be processed by computer vision algorithms, such as CNN-based methods [33], [34] and vision transformer-based methods [35], [36], [37], [38], to extract gait features for identification. For example, in [7], mmWave signals are analyzed in the range-Dopper domain using a CNN, achieving 92% accuracy in PID for up to four subjects.

Point cloud-based methods leverage advancements in radar devices, incorporating more antennas for collecting point clouds from human subjects. Recent developments in PID or re-identification (Re-ID) [8], [9], [10] showcase the use of recurrent neural networks (RNN), as seen in [8], achieving 89% PID accuracy for 12 subjects, even distinguishing two concurrently walking subjects.

To our knowledge, currently available mmWave radar point cloud datasets and PID methods lack consideration for occlusions and interference caused by obstacles in the scene. Consequently, our focus is on tackling mmWave radar-based PID under heavy occlusions. We achieve this by creating a new dataset and developing a suitable network model to overcome the challenges.

### B. Data Augmentation for Point Clouds

In scenarios with limited available training samples, employing data augmentation is a common strategy to mitigate overfitting and enhance model robustness by diversifying and expanding the sample set. Standard augmentation techniques for point cloud datasets include sample dropping, scaling, translation, rotation, and point-wise jittering [39], [40].

Recent advancements in point cloud augmentation methods are more sophisticated. PointMixUp [41] extends the mixup technique [42] from the image domain to point clouds. This method interpolates a new sample between two point cloud samples by finding the shortest path. PointAugment [43], an auto-augmentation framework, learns through adversarial training to generate samples aligning best with the classifier for point cloud classification. Additionally, PatchAugment [44] can be seamlessly integrated into the model for localized point cloud augmentation.

While most current methods emphasize augmenting point cloud datasets to improve static object classification, they may not effectively handle sparse radar point clouds. In this paper, we introduce a novel data augmentation approach tailored to mmWave radar point clouds. This method explicitly captures motion relationships to enhance the unique features of radar data, enriching its representation through a specialized data augmentation step.

### C. Open-Set Recognition (OSR)

Open-set recognition characterizes scenarios where new classes, unseen during training, emerge in testing. Classifiers are required to accurately identify known classes while refraining from assigning known labels to unseen classes [19]. Pioneering work in [45] establishes the foundations of the open-set recognition task. The initial deep learning approach for OSR,

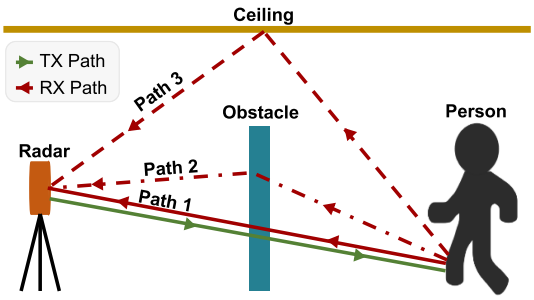



Fig. 3. Multipath propagation of mmWave radar signals in an occluded person identification scenario. The terms "TX Path" and "RX Path" refer to the transmission and reflection paths of radar signals between the radar sensor and the human subject, respectively.

named OpenMax, is introduced by [21], leveraging the Weibull model from the extreme value distribution. Additionally, generative adversarial networks (GANs) have been employed for this task [22], [46], with Neal et al. [46] generating a dataset using GANs, including unknown samples. This dataset, combined with a standard classification dataset, is then utilized to train an open-set classifier.

Other approaches include reconstruction-based methods [47], [48], [49], wherein poor test-time reconstruction serves as an open-set indicator. Prototype-based methods [50], [51] represent known classes through learned prototypes and identify open-set samples based on distances to these prototypes.

In the radar-based OSR task, Ni et al. [23] propose an approach utilizing a deep discriminative representation network and the Weibull model for open-set PID based on micro-Doppler signatures. Another study by [52] introduces an ensemble learning approach, incorporating reconstruction and multivariate Gaussian models, for the micro-Dopper-based open-set PID task. In contrast to these prior methods, our work aims to address the open-set PID problem using radar point clouds for heavy occlusions. Our experimental results demonstrate the efficacy of our method, achieving satisfactory performance on the open-set PID task.

## III. Overview and Problem Statement

In this section, we present an overview of the theoretical analysis and the problem statement.

### A. Mmwave Signals in Occlusion Scenarios

In this study, we use a frequency-modulated continuous wave (FMCW) mmWave radar to identify individuals in occluded scenarios, where the reflected signals from an individual propagate along multiple paths, as shown in Fig. 3. At the time slot $kT_c \leq t \leq (k+1)T_c$ at the $k$-th FMCW chirp, where $T_c$ denotes the duration of the chirp, the received intermediate frequency (IF) radar signal can be formulated as:

$$g(t) = g_D(t) + g_I(t) + b, \tag{1}$$

where $g_D(t)$ and $g_I(t)$ denote the signal components reflected by an individual from the direct and indirect paths, respectively, and $b$ represents the noise term, i.e., reflections from the environment other than the individual [53].

As shown by Path 1in Fig. 3, for the received signal component $g_D(t)$ reflected by an individual through a direct path, the IF signal can be represented as [53]:

$$g_D(t) = \frac{1}{2}\Gamma A_0 e^{i2\pi f_r(t-kT_c)} e^{-i2\pi f_v kT_c} e^{i4\pi f_0 r/\zeta}, \tag{2}$$

where $A_0$ and $f_0$ are the amplitude and carrier frequency of the transmitted signal, respectively. $r$ is the range between the subject and the receiver. $\zeta$ and $v$ respectively represent the speed of light and the relative velocity of the subject. $f_r = 2\varrho r/\zeta - 2vf_0/\zeta$ and $f_v = 2vf_0/\zeta$ respectively denote the fast and slow time frequencies. $\varrho = B/T_c$ is the slope of the sweep frequency, where $B$ is the sweep bandwidth. $\Gamma$ denotes the attenuation factor determined by the radar cross section (RCS) of the subject and propagation loss due to obstacle penetration and propagation distance. As reported by [16], [54], [55], the attenuation factor is affected by the materials and dimensions of obstacles, which can weaken the signal or even make it undetectable. We further analyze the impact of obstacles on mmWave radar signals in VI-E.

For the received signal component through an indirect path, the component undergoes additional reflections by the obstacle and other objects in the environment, as shown by Path 2 (by obstacle) and Path 3 (by ceiling) in Fig. 3. If we use $g_I^O$ and $g_I^E$ to represent the secondary reflection components by the obstacle and other environmental objects, $g_I(t)$ in (1) can be expressed as: $g_I(t) = g_I^O(t) + g_I^E(t)$.

For reflection components due to an obstacle, e.g., through Path 2, the IF radar signal can be formulated as [53]:

$$g_I^O(t) = g_D(t) \cdot \rho e^{i2\pi\Delta f_r(t-kT_c)} e^{-i2\pi\Delta f_v kT_c} e^{i4\pi f_0\Delta_r\zeta}, \tag{3}$$

where $\rho$ is the reflectivity of the obstacle surface responsible for the second reflection. $\Delta f_r \approx 2\varrho\Delta_r/\zeta$ and $\Delta f_v = 2\Delta_v f_0/\zeta$ represent the fast-time and slow-time frequency differences between signals from Path 1 and Path 2, respectively. $\Delta_r$ is the length difference between two propagation paths, and $\Delta_v$ is the radial velocity difference present in the radar signals. Note that due to the relay reflections from other objects than obstacles in the environment, the IF signal $g_I^E(t)$ in Path 3 has the same formulation as that of Path 2, but with a different attenuation factor, as it does not penetrate obstacles [56].

To summarize, as shown in (1), the radar IF signal experiences attenuation from the direct path and multipath interference from the indirect path. This results in power fading and a more complicated multipath point cloud, degrading person identification performance. However, multipath propagation can also provide valuable information for radar systems, such as human sensing through relay reflections [57], and wireless communication [58], [59].

### B. Problem Statement

The mmWave radar data that we use for person identification is the radar point cloud data, which can be calculated from the

IF signal in (1) through fast fourier transform (FFT) [60]:

$$P = \mathcal{F}(g), \tag{4}$$

where $\mathcal{F}(\cdot)$ denotes three distinct FFT operations, including Range-FFT, Doppler-FFT and Angle-FFT, for measuring the ranges, velocities and angles of the target. Each radar point cloud $P$ contains multiple points consisting of four attributes $p_n = (x_n, y_n, z_n, v_n), n \in [1, N]$, where $N$ is the total number of points in the point cloud. $(x_n, y_n, z_n)$ represents 3D coordinates of the point, and $v_n$ denotes the velocity.

The point cloud $P$ obtained from different individuals contains distinct gait information used for identification. However, due to the interference from the obstacle and sparse nature of mmWave point clouds, it is insufficient to use a single point cloud for person identification. Thus, we consider using a sequence of radar point clouds $S$ for this purpose. Given a classifier $\Omega$, the point cloud sequence $S$ is input into $\Omega$, which outputs the probabilities associated with each identity. This can be formulated as:

$$\mathcal{P}(c|S) = \Omega(S), \tag{5}$$

where $c \in \{1, 2, \ldots, C\}$ represents the $c$-th identity, and $\mathcal{P}(c|S)$ denotes the probability of the sample $S$ belonging to the $c$-th identity.

Furthermore, for open-set person identification, we need an open-set classifier $\Lambda$ that can not only identify the target from known classes but also recognize unknown classes. It is defined as:

$$\mathcal{P}(c|S) = \Lambda(S) \tag{6}$$

where $c \in \{1, 2, \ldots, C, C+1\}$ represents the $c$-th identity class, with $C+1$ denoting the unknown identity class.

In this study, we design a novel classifier $\Omega$ using neural networks and supervised contrastive learning. Furthermore, we propose combining Weibull models with the pre-trained $\Omega$ as the classifier $\Lambda$ for open-set person identification.

## IV. Method

Our OSO-PID framework includes two major components: the mmWave point cloud-based PID described in Section IV-A and the open-set recognition described in Section IV-B.

### A. mmWave Radar Person Identification

For person identification under occlusions, we first propose to use a novel data augmentation strategy, called **point motion signature (PMS)** to take advantage of the spatial-temporal relationships between consecutive point cloud frames. As a result, mm-PIDNet uses a sequence of augmented point clouds as input. Then, the proposed mm-PIDNet uses a content encoder consisting of fully connected layers, which generates a high-dimensional feature for each point cloud in the sequence. After that, we use a transformer encoder with a self-attention mechanism to learn the spatial relationships within the sequence of features. The output from the transformer encoder is fed into a Bi-LSTM network for learning the temporal relationships, and the last hidden states of both forward and backward passes of Bi-LSTM are summed to produce a global vector. The vector is used for identifying the individual through linear layers and a softmax layer. Simultaneously, we propose a supervised contrastive learning module to minimize intra-class distance and maximize inter-class distance in the feature space. To this end, we implement an offline buffer for the global vectors from Bi-LSTM and use the momentum update method [27] to maintain feature consistency between the vectors stored in the buffer and those currently predicted by the network. The contrastive loss is used to update the network to learn discriminative features for each individual. Fig. 4 illustrates the proposed pipeline and architecture. Details are presented in the following.

*1) Data Augmentation:* Due to attenuation caused by obstacles, mmWave point clouds are sparser, leading to potential performance degradation in subject identification. To overcome this challenge, we introduce a data augmentation strategy designed to leverage the spatial-temporal relationships among radar frames to enhance PID performance.

We propose a new data augmentation strategy named point motion signature (PMS) to accurately capture the dynamic changes occurring between adjacent radar frames. Our data augmentation is based on interpolation analysis across both spatial and velocity variations, to gather a richer set of motion information. This involves utilizing the current point's attributes to calculate the differences with the attributes of the three nearest points in the preceding point cloud. More specifically, in the context of two adjacent radar point cloud frames $P_l$ and $P_{l+1}$, we initially choose a point from $P_{l+1}$. Subsequently, we calculate the three closest points in $P_l$ based on euclidean distance. The differences in coordinates and the velocity are merged with the initial point signatures via concatenation.

Our design is partly inspired by the point flow method [61], which selects the closest points in adjacent frames and calculates the velocity difference of the two nearest points as a new signature. However, solely focusing on velocity disparity fails to adequately capture the inter-frame motion relationship. Taking spatial differences into account, our method not only increases the training data volume by 4 times compared to its original size but also provides a richer set of dynamic information related to human motion. Experiments show that it greatly improves PID performance.

*2) Network Structure for Person Identification:* As illustrated in Fig. 4, the proposed PIDNet comprises three main modules. The content encoder maps the augmented point cloud sequence into a high-dimensional feature space and subsequently extracts a high-dimensional feature from each point cloud in the sequence. The spatial-temporal extractor incorporates a transformer encoder and a Bi-LSTM module to capture the spatial-temporal relationships of successive point cloud frames. Additionally, supervised contrastive learning establishes an offline buffer designed to store features generated by Bi-LSTM, which are used to compute the contrastive loss. Finally, linear layers are used with a softmax function to determine the person's identity.

*Content encoder:* Given a sequence of point cloud frames, the content encoder takes each point cloud $P_l$ with $N$ points as input. Each point $p_n^l$ ($n \in [1, N]$) encompasses sixteen attributes, including the 3D coordinates, velocity, and PMS signatures.

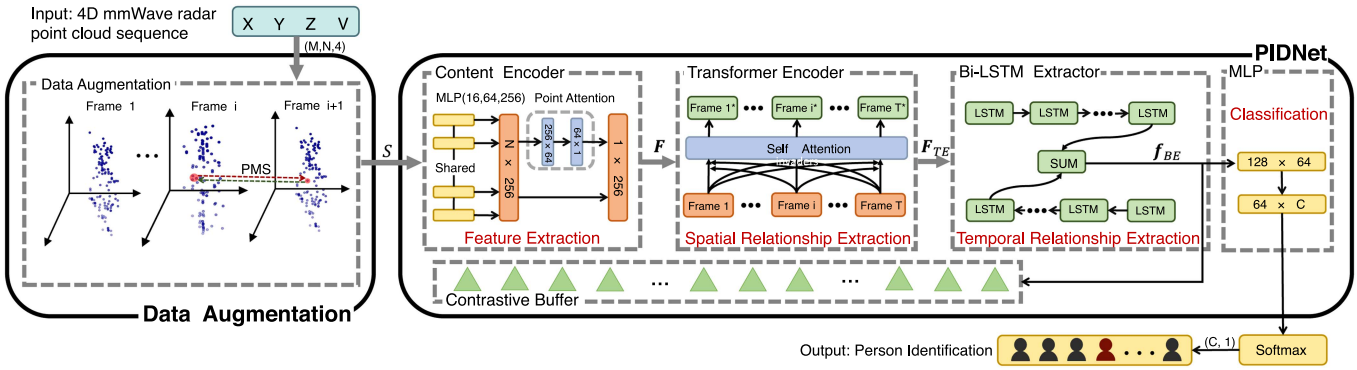



Fig. 4. Architecture of our mmWave radar-based PID model with two major components: data augmentation and PIDNet. The data augmentation module enriches the motion features and outputs a sequence of augmented point clouds $S$ for identification improvement. The PIDNet model takes $S$ as input and consists of three components for feature extraction: (1) a content encoder, comprising MLP layers, which extracts features from each point cloud with an attention function and outputs a sequence of high-dimensional features, $\mathbf{F}$; (2) a transformer encoder, which takes $\mathbf{F}$ as input to enriches spatial semantic information in the features; and (3) a Bi-LSTM encoder, which takes the transformer encoder's output, $\mathbf{F}_{TE}$, as input to capture temporal relationships and generate a comprehensive gait feature, $\mathbf{f}_{BE}$. The final MLP processes $\mathbf{f}_{BE}$ to produce the identification result.

The content encoder processes each point independently using a shared-weight multi-layer perceptron (MLP) $\mathcal{M}$ and produces a high-level feature denoted as $\mathbf{f}_n^l = \mathcal{M}(p_n^l; \Theta_{\mathcal{M}})$, where $\Theta_{\mathcal{M}}$ represents the learnable parameters of $\mathcal{M}$. Subsequently, the features of all points within a frame are aggregated into a global feature using a point attention function. This function assigns a score to each point in the present point cloud and calculates a weighted sum of all points. This adaptive mechanism allows the feature extractor to adjust the contribution of each point based on the significance of identity-specific features associated with the corresponding body part. The attention procedure is computed as follows:

$$\mathbf{f}_l = \sum_{n=1}^{N} \mathcal{A}_p\left(\mathbf{f}_n^l; \Theta_{\mathcal{A}_p}\right) \times \mathbf{f}_n^l, \tag{7}$$

where $\mathcal{A}_p(\cdot)$ is the point attention implemented by two linear layers and a softmax function, $\Theta_{\mathcal{A}_p}$ is the set of parameters belonging to the point attention. The two linear layers comprise 256 and 64 hidden units, respectively.

*Transformer encoder:* Given the sparse nature of radar point clouds, the inherent attributes within a single frame are insufficient for effective learning. Therefore, enhancing spatial features for each point cloud frame by leveraging the input point cloud sequence becomes crucial. Drawing inspiration from the transformer architecture [25], originally designed for processing word sequences, we employ a transformer encoder to boost spatial semantic information for each point cloud.

The self-attention mechanism [62] serves as the fundamental element within this layer, generating refined attention features based on the input point cloud sequence. In particular, it takes a sequence of global features of point clouds as input and computes three vectors for each feature: *query*, *key* and *value* through linear transformations. Three matrices $\mathbf{Q} \in \mathbb{R}^{L \times d_a}$, $\mathbf{K} \in \mathbb{R}^{L \times d_a}$ and $\mathbf{V} \in \mathbb{R}^{L \times d_e}$ are the results of linear transformations applied to the input feature matrix $\mathbf{F} \in \mathbb{R}^{L \times d_e}$, respectively, where $L$ denotes the length of the input sequence, $d_e$ represents the dimension of the global feature of each point cloud, and $d_a$ denotes the dimension of the query and key vectors. The calculation is expressed as:

$$(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{F} \cdot (\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v), \tag{8}$$

where $\mathbf{W}_q \in \mathbb{R}^{d_e \times d_a}$, $\mathbf{W}_k \in \mathbb{R}^{d_e \times d_a}$ and $\mathbf{W}_v \in \mathbb{R}^{d_e \times d_a}$ are learnable parameter matrices. In this work, we set $d_a$ to be $d_e/4$ for computational efficiency. Subsequently, $\mathbf{Q}$ and $\mathbf{K}$ are employed to compute the softmax normalized attention weights using matrix dot-product operations.

$$\mathbf{A} = \mathcal{S}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_a}}\right), \tag{9}$$

where $\mathcal{S}$ represents the softmax function. $\mathbf{A}$ is a weight matrix used to multiply the *value* matrix $\mathbf{V}$ to produce self-attention features. These features are subsequently normalized through a layer normalization function $\mathcal{L}$, followed by a residual connection operation:

$$\mathbf{F}_O = \mathcal{L}(\mathbf{A}\mathbf{V}) + \mathbf{F}. \tag{10}$$

The resulting feature matrix $\mathbf{F}_O$ is fed to a feed-forward neural network $\mathcal{N}$, followed by another layer normalization and residual connection to generate the final output $\mathbf{F}_{TE} = \mathcal{L}(\mathcal{N}(\mathbf{F}_O)) + \mathbf{F}_O$.

For each point cloud global feature of the input sequence, the self-attention mechanism is employed to calculate dynamic weights to all features within the matrix $\mathbf{F}$. Subsequently, the weight matrix $\mathbf{A}$ will be utilized to conduct a weighted summation of the value matrix $\mathbf{V}$, thereby significantly enhancing the infusion of semantic information in the derivation of $\mathbf{F}_{TE}$. Moreover, the inclusion of the layer normalization $\mathcal{L}$ along with the residual connection serves to bolster stability and effectively mitigate the challenge of vanishing gradients.

*Bi-LSTM gait feature extractor:* Due to the sparsity of radar point clouds, the gait information in a single radar point cloud is incomplete. This indicates a single high-dimensional feature from a radar point cloud is insufficient to characterize a target person. The transformer layer augments more spatial information within each radar point cloud feature present in the current input sequence. We further feed the sequence of the extracted features $\mathbf{F}_{TE} = [\mathbf{f}_1^{TE}, \mathbf{f}_2^{TE}, \ldots, \mathbf{f}_L^{TE}]$ (where $L$

is the length of the sequence) into Bi-LSTM [26], obtaining a comprehensive gait feature across the temporal dimension to improve the identification performance.

Following the input of the adapted features into Bi-LSTM, the last hidden state of the forward LSTM and the last hidden state of the backward LSTM are summed to represent the global gait feature. Additionally, two fully connected layers and a softmax function are used to produce the output of the classification probabilities for $C$ classes.

*3) Supervised Contrastive Learning:* In the realm of effective representation learning, contrastive learning [27] guides pairs of samples to approach or diverge in feature space based on their class labels during the training process. In line with this principle, we introduce a supervised contrastive learning method to encourage samples with the same class labels to be closer in feature space, while those with distinct labels experience greater separation. This method serves a dual purpose: (1) mitigating model overfitting in the PID task, and (2) crafting a compact feature space suitable for addressing open-set problems.

Due to the GPU storage constraints, traditional contrastive learning methods often focus on assessing similarity within the current input batch. In our approach, we introduce an offline buffer specifically designed for contrastive samples, illustrated in Fig. 4 bottom. This buffer operates as a data queue with a capacity that surpasses the limitation of a standard mini-batch size. In our experiments, the buffer size is configured as 8,192.

During training, samples in the buffer are continuously refreshed. Features generated by the Bi-LSTM extractor are fed into the offline buffer. When the quantity of samples in the buffer surpasses its capacity, the oldest mini-batch sample will be removed. This operation is advantageous as the oldest sample tends to be the most outdated and least consistent with the latest ones. Moreover, to further maintain the consistency between the samples in the buffer and the current mini-batch sample, we employ the momentum update method introduced by [27]. In the training phase, the PID network without final fully connected layers as the encoder $\mathcal{Q}$, generates the newest sample to replace the oldest ones in the buffer. Simultaneously, an additional encoder $\mathcal{G}$ sharing the same structure as encoder $\mathcal{Q}$, is introduced to update all samples in the buffer. Formally, the parameters of $\mathcal{Q}$ and $\mathcal{G}$ are denoted as $\Theta_{\mathcal{Q}}$ and $\Theta_{\mathcal{G}}$, respectively. The update of $\Theta_{\mathcal{G}}$ is performed by:

$$\Theta_{\mathcal{G}} = \sigma\Theta_{\mathcal{G}} + (1-\sigma)\Theta_{\mathcal{Q}}, \tag{11}$$

where $\sigma \in [0, 1)$ is a momentum coefficient. During the learning process, back-propagation updates only the parameters $\Theta_{\mathcal{Q}}$, while the update of the encoder $\mathcal{G}$ relies on the parameters $\Theta_{\mathcal{Q}}$. The momentum update ensures a smoother evolution of $\Theta_{\mathcal{G}}$ compared to $\Theta_{\mathcal{Q}}$ and maintains the coherence of samples between the buffer and the current mini-batch.

*4) Loss Functions:* In the learning process, the PID network is primarily optimized using a combination of the cross-entropy loss and the contrastive loss. Let $M$ denote the number of training samples in the current batch, $y_{mc}$ represents the ground truth label, taking the value of 1 if the sample belongs to $c$-th person and 0 otherwise. The symbol $\hat{y}_{mc}$ denotes the probability that the sample belongs to $c$-th ID class. The cross-entropy loss is formulated as:

$$L_{cro} = -\frac{1}{M}\sum_{m}^{M}\sum_{c}^{C} y_{mc}\ \log\left(\hat{y}_{mc}\right). \tag{12}$$

For the contrastive learning loss, we compute the similarity between the samples in the current batch and the samples updated in the buffer using the encoder $\mathcal{G}$ via the mean squared error (MSE) loss. Let $U$ denote the buffer size, and $\mathcal{D}()$ denote the similarity function quantifying the relationship between a feature $\mathbf{f}_{ba}$ from the current batch and a contrastive feature $\mathbf{f}_{bu}$ from the buffer, calculated via a dot product. Let $l_o \in \{-1, 1\}$ denote the pre-defined similarity label determined by whether $\mathbf{f}_{ba}$ and $\mathbf{f}_{bu}$ share the same class label or not. The MSE loss is calculated as:

$$L_{mse} = \frac{1}{MU}\sum_{m}^{M}\sum_{u}^{U} \left\|\mathcal{D}\left(\mathbf{f}_{ba}, \mathbf{f}_{bu}\right), l_o\right\|^2. \tag{13}$$

The final loss function for the person identification is:

$$L_{ide} = L_{cro} + \epsilon\, L_{mse}, \tag{14}$$

where $\epsilon$ is a hyper-parameter determining the relative influence of outcomes from the two distinct loss functions. In our experimental setup, $\epsilon = 0.1$.

In our experiment, the cross-entropy loss is pivotal in determining the performance of person identification. Simultaneously, the supervised contrastive learning loss serves the dual purposes of mitigating model overfitting and establishing a compact feature space to address the open-set PID problem.

### B. Open-Set Recognition

For open-set person identification, as depicted in Fig. 5, we first establish a pre-trained mm-PIDNet using training samples from known classes. During mm-PIDNet training, our supervised contrastive learning method draws features with the same identity label closer together while pushing those with different identity labels farther apart, resulting in a compact feature space for each individual. When a sample from an unknown person is fed into the pre-trained mm-PIDNet, the high-dimensional feature may not fall into the feature region of each known individual. Thus, we propose to use the Weibull model, which has an upper-bounded distribution function, to further refine the support region of each known class in the feature space. The features of correctly identified training samples, extracted from mm-PIDNet before the softmax layer, are used as "activation vectors". These vectors are averaged to compute the class center for each individual. Then, we compute cosine distances between activation vectors and the corresponding class center, and the largest 20 distances are used to fit the Weibull model for each class. When a testing sample is input into mm-PIDNet, the probabilities belonging to each known identity and the unknown identity are determined by the fitted Weibull models.

*1) The Closed-Set Pre-Trained PID Model:* To effectively utilize latent representations, we pre-train a mm-PIDNet model under the closed-set classification assumption, as shown in Fig. 5. Initially, the PID model is trained using samples with known ID classes, optimized under the identification loss $L_{ide}$

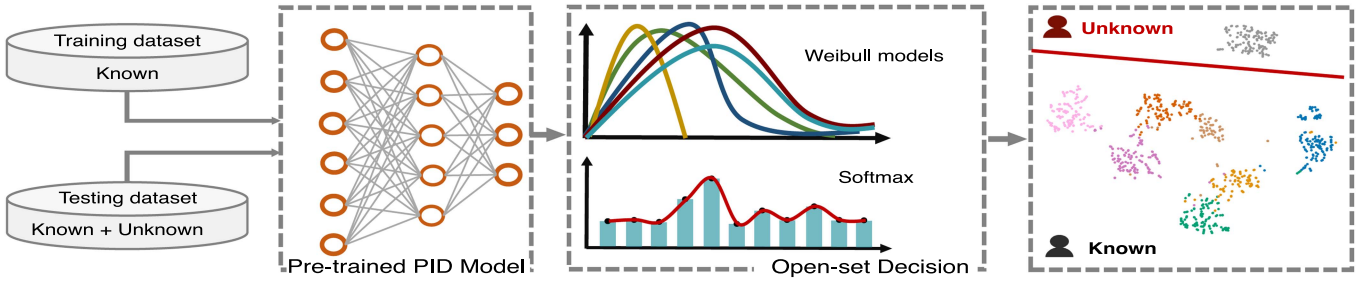



Fig. 5. The open-set occluded person identification method consists of two modules: (1) a pre-trained mm-PIDNet model for extracting gait features and computing the distance with the corresponding *class centers*, and (2) a second module that fits Weibull models for determining open-set classification using the obtained distance measures.

and supervised contrastive learning loss $L_{mse}$. Subsequently, the softmax layer of the pre-trained mm-PIDNet (as indicated in Fig. 4) is removed. The output of pre-trained mm-PIDNet will be employed for fitting Weibull models in subsequent steps.

*2) Fitting Weibull Models:* We apply the extreme value theory (EVT) to open-set recognition, where EVT provides a functional form for modeling the probability of which a sample belongs to each class, including the unknown class [21], [63]. According to the Fisher-Tippett theorem [64], EVT states that the distribution of extreme values from a sequence of independent and identically distributed (*i.i.d.*) random variables converges to one of three possible forms: the Gumbel distribution, Frechet distribution, or reversed Weibull distribution. While the Gumbel and Frechet distributions are suited for unbounded data, the Weibull distribution is used for bounded data [65]. Previous studies [66], [67] have shown that the open-set samples follow a Weibull distribution due to their bounded nature. Therefore, we use the Weibull distribution in our framework to calculate the probabilities that a sample belongs to each identity, including the unknown one.

For each known identity, we first compute the class center using correctly identified training samples. These samples are input into a pre-trained mm-PIDNet without the softmax layer, and the resulting outputs are averaged to generate the class center. Then, we compute the cosine distances between features of correctly classified training samples and their corresponding class centers, yielding the class-specific distance distribution for each known identity. After that, we fit a class-specific Weibull model using the 20 largest cosine distances of each class. The cumulative distribution function (CDF) of the fitter Weibull model for the $c$-th class with correctly classified training samples is formulated as [68]:

$$\mathcal{W}_c(d_c;\lambda_c,\beta_c)=\begin{cases}1-\exp\left[-\left(\frac{d_c}{\lambda_c}\right)^{\beta_c}\right] & \text{if } d_c\geq 0,\\ 0, & \text{otherwise,}\end{cases} \tag{15}$$

where $\lambda_c > 0$ and $\beta_c > 0$ are the scale and shape parameters, respectively, and $d_c$ represents the cosine distance between the feature vector of an input sequence of point clouds and the $c$-th class center. Parameters of the Weibull model $(\lambda_c, \beta_c)$ can be estimated by the procedure described in [69].

For a point cloud sequence sample $S$, the open-set score $\hat{\Psi}_c$ for the feature vector $\boldsymbol{\nu} = \text{mm-PIDNet}(S)$ belonging to the $c$-th known class is calculated as:

$$\hat{\Psi}_c\left(d_c;\lambda_c,\beta_c\right)=\boldsymbol{\nu}_c\cdot\left[1-\frac{\gamma-c}{\gamma}\mathcal{W}_c\left(d_c;\lambda_c,\beta_c\right)\right], \tag{16}$$

where $\boldsymbol{\nu}_c$ is the $c$-th element of the vector $\boldsymbol{\nu}$. $\gamma$ denotes the number of top classes to be revised that are sorted by distance, with $\gamma = 7$ by default.

According to (15), when the parameters $\lambda_c$ and $\beta_c$ are provided, $\mathcal{W}(d_c;\lambda_c,\beta_c)$ exhibits a monotonically increasing pattern as $d_c$ increases, consequently causing a decrease in the open-set score $\hat{\Psi}_c$. This suggests that the likelihood of a sample belonging to a known class decreases as the distance from the sample to the class center increases. Then, the open-set score $\hat{\Psi}_{C+1}$ indicating a sample belonging to the unknown class can be calculated as:

$$\hat{\Psi}_{C+1}=1-\sum_{c}^{C}\hat{\Psi}_c\left(d_c;\lambda_c,\beta_c\right), \tag{17}$$

where $C$ is the total number of known identity classes.

Lastly, the vector of probabilities of the input sequence $S$ belonging to each of the identity $c$ in $C+1$ classes (including the unknown identity class) can be computed by the softmax function:

$$\mathcal{P}(c|S)=\frac{e^{\hat{\Psi}_c}}{\sum_{c=1}^{c=C+1}e^{\hat{\Psi}_c}}. \tag{18}$$

*3) Open-Set PID Decision Making:* For a given point cloud sequence sample of a person in query and $C$ number of known identities, the open-set PID task is to determine the person's identity among the $C+1$ classes, where the $(C+1)$-th label represents the unknown (intruder) class.

In the testing phase, the open-set identification model predicts the identity label $\hat{y}$ for the point cloud sequence sample $S$ of a person in query. Let $\mathcal{P}(c|S)$ denote the probability belonging to the $c$-th identity computed by (18). We use a threshold $\delta$ default to 0.4 to determine if the sample $S$ belongs to one of the known classes. This decision function assigns a known identity label if the highest probability exceeds threshold $\delta$; otherwise, it rejects the input as an unknown intruder:

$$\hat{y}=\begin{cases}\underset{c\in\{1,\ldots,C+1\}}{\arg\max}\ \mathcal{P}(c|S), & \text{if } \mathcal{P}(c|S)\geq\delta,\\ \text{unknown intruder ID}, & \text{otherwise.}\end{cases} \tag{19}$$

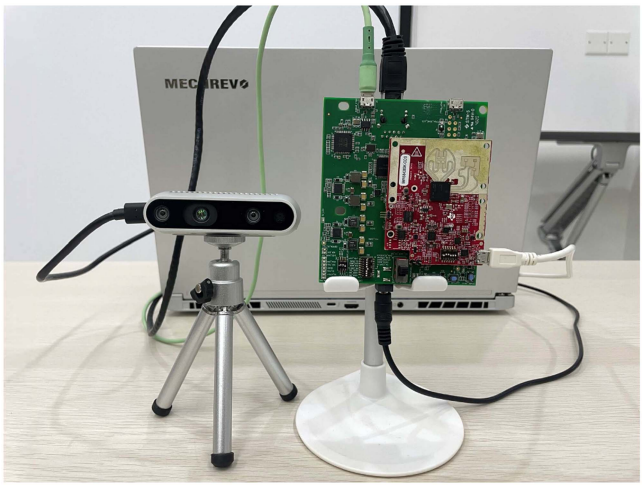


Fig. 6. PID sensing devices: an RGB-D camera for capturing RGB and depth images, and a mmWave radar with a DCA1000EVM data collection card collected to a laptop.

## V. Experiment and Dataset

This section describes our experimental setup as well as the newly collected **mmWave-ocPID** dataset for the training and evaluation of mmWave PID models under heavy occlusions.

### A. Experimental Setup

We employ a COTS FMCW radar along with an RGB-D camera to capture sequential data of individuals walking behind obstacles, as illustrated in Fig. 6. We use an indoor environment featuring multiple configurations of layouts and obstacles, as depicted in Fig. 2(a).

*MmWave radar and camera:* We use a TI IWR6843ISK-ODS radar for transmitting and receiving radar signals, paired with an Intel RealSense D435 camera for capturing RGB and depth images at the resolution of $424 \times 240$.

The radar configuration includes three transmitter antennas and four receiver antennas, transmitting 128 chirps per frame and sampling 256 signals per chirp. The radar settings are as follows: start frequency at 60.05 GHz, frequency slope at 130.029 MHz/$\mu$s, and bandwidth at 3.9 GHz. We set the chirp cycle time to 30 $\mu$s, idle time to 100 $\mu$s, and sampling start time as 2 $\mu$s. This configuration yields a range resolution of 4.5 cm and a maximum unambiguous range of 11.52 m. It allows measurement of a maximum radial velocity of 3 m/s with a resolution of 0.049m/s. Radar frames are sampled at 15 Hz, while RGB and depth images are acquired at 30 Hz. Raw signals from the radar to a laptop are transmitted using a DCA1000EVM data capture card.

*Obstacles:* Considering the diverse electromagnetic wave absorption and reflection properties of different materials, our experiment employs three types of obstacles: a clothes rack, a poster board, and a potted plant. The radar and camera devices are positioned on a table at 0.9 m height, while the obstacles are placed 1.5 m away. The clothes rack stands at 1.5 m height, the poster board is 1.8 m tall, and the pot plant has an approximate height of 1.5 m within a 0.5 m ceramic flower pot. To enhance realism, three pieces of clothing are hung on the designated rack. To broaden the scope of scenarios and comprehensively evaluate radar sensor performance, walking is conducted in the indoor environment without occlusion.

*Occlusions:* In our experiments, we use three types of obstacles: a clothes rack with three pieces of clothing, a poster board, and a potted plant. We choose these obstacles because they are common objects encountered in real-world scenarios, and they contain different materials and dimensions, producing different attenuation and multipath effects on mmWave radar signals. In addition, three types of occlusions are observed during data collection. First, the poster board completely blocks the line of sight of both the radar and camera, resulting in full occlusion of the target. Second, when the clothes rack and potted plant are used as obstacles, some signals can pass through gaps in clothing and plant leaves, leading to heavy occlusion of the target. Third, we also collect multi-modal data from the same experimental configuration without occlusion for comprehensive evaluations and comparisons.

*Volunteers:* Our dataset comprises samples collected from 23 individuals, including 20 males and 3 females. The recruited volunteers have ages ranging from 22 to 41 years old, heights from 1.65 m to 1.90 m, and weights from 50 Kg to 100 Kg. In each occlusion scenario, participants are instructed to perform inbound/outbound walking behind the obstacles. Each subject completes 4 consecutive minutes of walking, capturing 3,600 radar frames and 7,200 RGB and depth images by the devices.

*Signal processing:* We first use the FFT [60] and CA-CFAR [70] algorithms to generate radar point clouds for evaluating the OSO-PID framework. It includes the use of Range-FFT and Doppler-FFT on raw signal data, followed by CA-CFAR on the range-Dopper map to detect peak value indices. These indices guide Angle-FFT to produce point clouds. To diversify and thoroughly assess the CA-CFAR point clouds, we experiment with false alarm rates (FAR) of 0.02, 0.05, and 0.08. Meanwhile, we use the maximum-energy time-frequency ridge extraction method [71] to generate radar point clouds, which locates the time-frequency ridge in the range profile to determine subject range bins in the range-Dopper map. Considering the non-rigid human body, we select 10 range bins near the ridge to represent various body segments. We then choose the top 256 indices with the highest energy from the associated region in the range-Dopper map for further processing via Angle-FFT. Our experiment sets a penalty factor of 0.05 for the maximum-energy time-frequency ridge method. In total, we construct two types of radar point cloud datasets using the CA-CFAR and the time-frequency ridge algorithms, respectively.

### B. mmWave-OcPID Dataset

The dataset contains radar point clouds, RGB and depth images, as depicted in Fig. 2. In each scenario, 23 subjects individually engage in a 4-minute walking session, resulting in the capture of 3,600 frames per person. Within each scenario, approximately 80,000 frames are collected. Our experiment has 4 scenarios, where each represents a distinct occlusion condition. Despite minimal data loss, the mmWave-ocPID dataset contains

over 300,000 frames of mmWave radar measurements and over 600,000 RGB and depth images.

In the radar signal collection process, inherent constraints such as the limited number of antennas and poor angular resolution result in sparse point cloud capture. After employing signal processing, some point clouds may still be noisy. To this end, we use a heuristic method to constrain the $X$ dimension and velocity $V$ to $\pm 1\,\mathrm{m}$ and $\pm 2\,\mathrm{m/s}$, and set the minimum $Z$ dimension value to $-0.9\,\mathrm{m}$. These heuristic rules improve overall quality, however reduce the number of points associated with each individual in each frame. To address this issue, the data augmentation method from Section IV-A is performed to enrich the point cloud samples.

Compared with other mmWave radar-based PID datasets, e.g., the mmGait dataset [9], our mmWave-ocPID dataset is a multi-modal dataset comprising radar data, RGB, and depth images with a variety of occlusion scenarios. We integrate various signal processing methods to generate point clouds, thereby augmenting the diversity of the dataset and facilitating comprehensive evaluations. We have obtained institutional review board (IRB) approval for all experiments conducted in this work, and we make the dataset publicly available in IEEE Dataport: https://dx.doi.org/10.21227/vkx6-fy49.

## VI. Evaluations

### A. Implement Details

*1) Person Identification Settings:* In our experimental setup, each input point cloud sequence sample contains 45 point clouds. Note that, the content encoder network processes each point cloud individually, taking a fixed dimensional of 256 points as input. After applying the data acquisition constraints outlined in Section V-B, certain point clouds may contain fewer than 256 points. Hence, we augment the acquired point cloud by repeating points from the same cloud to pad it up to 256. Additionally, because the initial point cloud in each input sequence sample lacks PMS features, we zero-pad all points in the initial point cloud for both training and testing samples.

To ensure consistency across all environmental configurations during the dataset split, we maintain a fixed ratio of training frames to testing frames for each individual in each scenario, set at $7:1$. The samples are then split using a sliding window approach, with each window overlapping the preceding one by 40 frames.

Our PID network is trained end-to-end using the ADAM optimizer with a weight decay of $1e-5$. The initial learning rate is 0.001, and the batch size is fixed at 256. Training is conducted using an RTX 3090 GPU. The model incorporates shared-weight linear layers with input sizes of 16, 64, and 256, each followed by a ReLU activation function. The Bi-LSTM is configured with an input size of 256 and a hidden size of 128. The two classification layers have input sizes of 128 and 64, respectively, with the output of the first layer activated by a ReLU function.

*2) Open-Set Settings:* Open-set problems exhibit varying difficulty based on the ratios of the known to unknown categories in the testing set. Our approach aligns with the conceptual framework proposed by Geng et al. [19]. Let $N_{train}$ denote the number of training categories, and $N_{test}$ denote the number of testing categories. The **openness** is defined as:

$$openness = 1 - \sqrt{\frac{2N_{train}}{N_{test} + N_{train}}}, \tag{20}$$

The concept of openness is quantified by a value between 0 and 1. When $N_{train}$ equals $N_{test}$, openness is zero, reducing to a closed-set problem. The difficulty of the open-set task grows as the openness value increases. To demonstrate the algorithm's robustness across various openness levels, we create nine degrees of openness by randomly designating 11 classes as known and the remaining 12 as unknown, chosen from a pool of 23 total categories. Table I provides our experimental setup regarding the openness levels. Furthermore, we use the F1-score described in [52] to evaluate the open-set classification performance. The F1-score ranges from [0,1], and a higher value indicates the better performance of an open-set classifier.

TABLE I
Openness (%) Configurations in Our Experiments

| $N_{train}$ | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
|---|---|---|---|---|---|---|---|---|---|
| $N_{test}$ | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| Openness | 6.19 | 8.01 | 9.73 | 11.36 | 12.90 | 14.36 | 15.76 | 17.08 | 18.35 |

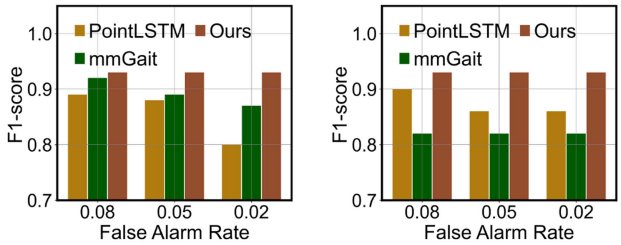
TABLE II
Average PID Performance Under Occlusion Scenarios, Evaluated Using F1-Scores at Fixed False Alarm Rates (FAR) and Time-Frequency Processing

| Method \ Dataset | CFAR FAR:0.08 | CFAR FAR:0.05 | CFAR FAR:0.02 | Time-Frequency Ridge |
|---|---|---|---|---|
| ResNet [33] | 0.89 | 0.88 | 0.80 | 0.90 |
| DenseNet [34] | 0.89 | 0.87 | 0.80 | 0.90 |
| Nest [38] | 0.91 | 0.87 | 0.82 | 0.88 |
| Vit [35] | 0.83 | 0.78 | 0.71 | 0.81 |
| PiT [36] | <u>0.92</u> | <u>0.89</u> | <u>0.87</u> | <u>0.91</u> |
| MobileViT [37] | 0.89 | 0.85 | 0.77 | 0.87 |
| MultiBranch [11] | 0.87 | 0.82 | 0.75 | 0.89 |
| MuID [7] | 0.84 | 0.79 | 0.61 | 0.78 |
| PointLSTM [10] | 0.90 | 0.86 | 0.82 | 0.84 |
| mmGait [9] | 0.82 | 0.82 | 0.82 | 0.76 |
| **Ours** | **0.93** | **0.93** | **0.93** | **0.92** |

The best and second-best results are marked in bold and underlined.

### B. Person Identification Performance

*1) Overall Performance:* Table II shows a comparison of average F1-scores between our approach and state-of-the-art methods in all scenarios. The compared methods fall into two categories: those using Doppler images and those using point clouds for person identification. To assess the effectiveness of Doppler images, we compare our method with six baseline models: ResNet [33], DenseNet [34], Nest [38], ViT [35], PiT [36] and MobileViT [37]. ResNet and DenseNet are standard CNN architectures commonly used in computer vision tasks, while the vision transformer-based models represent newer approaches




(a) Comparison between Dopper image-based methods and ours. (b) Comparison among point cloud-based methods.

Fig. 7. Comparison of PID accuracies at different false alarm rates.

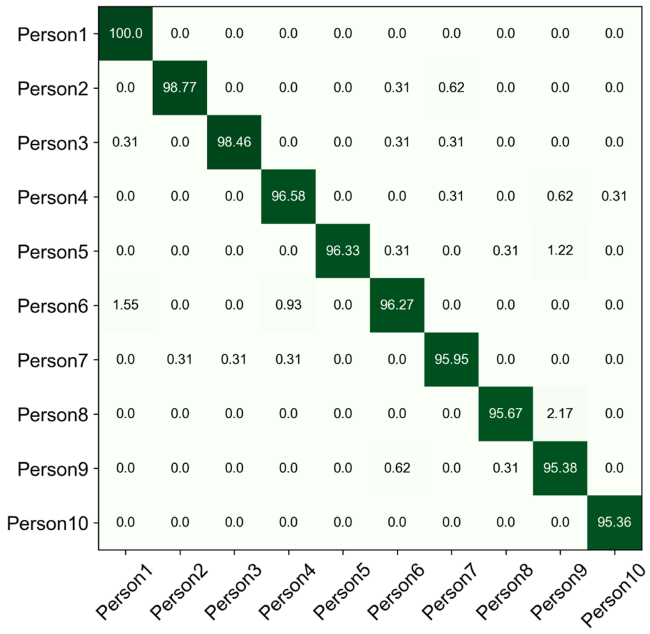
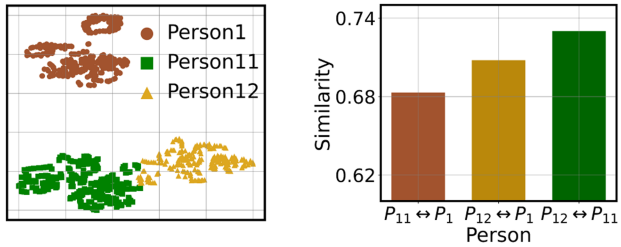



Fig. 8. The PID accuracy confusion matrix(%). Average accuracy scores greater than 95.36% are achieved for all subjects.

that have shown strong performance across various tasks. We evaluate four vision transformer-based methods [35], [36], [37], [38] that use Doppler images for person identification.

The results in Table II show that our method outperforms all compared methods, achieving F1-score values of 0.93, 0.93, 0.93 and 0.92 on the four datasets. Fig. 7(a) shows that the PID performance decreases as the false alarm rate is reduced. Our approach shows greater robustness in identification compared to alternative methods such as PiT [36] and ResNet [33]. This highlights that our method is well-suited for radar point clouds and does not require extensive parameter tuning in the signal processing algorithm. Furthermore, our approach is the first radar point cloud-based solution specifically designed for person identification in occlusion scenarios, unlike PointLSTM [10] and mmGait [9], which were designed for non-occlusion cases. As shown in Fig. 7(b), our method outperforms mmGait by at least 13.41% on various false alarm rates.

*2) Confusion Matrix:* Fig. 8 illustrates the percentage confusion matrix of identification results. The diagonal of the matrix displays the classification accuracy for the top 10 individuals with the highest identification performance. According to these




(a) Feature visualization (b) Average cosine similarity

Fig. 9. (a) Visualization of features collected from three individuals: Person 1, Person 11 and Person 12. (b) Average cosine similarities between features obtained from different individuals. $P_1$, $P_{11}$ and $P_{12}$ correspond to Person 1, Person 11 and Person 12, respectively.

results, all individual accuracies are above 95.36%, while 7 of them exceed 96.00%. We count the identification accuracy of all individuals, and over 90% of the total subjects achieve an identification accuracy exceeding 87%. These results demonstrate that mmWave radar measurements are sufficiently sensitive to capture significant gait features for PID under various occlusions. Additionally, we observe some people whose identification accuracies are below 90%. For example, the lowest identification accuracy in our experiment is 81.17% for Person 12, attributed to the occasional misidentification of Person 12 as Person 11. We use t-SNE [72] to visualize features extracted from Bi-LSTM for three individuals: Person 1, Person 11 and Person 12. As shown in Fig. 9(a), the visualization reveals that some features of Person 12 overlap with those of Person 11, while remaining more distant from the features of Person 1. This suggests that the features of Person 12 are more likely to be misidentified as those of Person 11, potentially reducing the identification accuracy for Person 12. Meanwhile, we use the cosine similarity metric to quantitatively assess the average feature similarity among the three individuals. As shown in Fig. 9(b), the average cosine similarity between Person 12 and Person 11 is higher than that between Person 12 and Person 1, as well as Person 11 and Person 1. This also verifies that the features of Person 12 are more prone to being misclassified as those of Person 11. These results suggest that our method has potential for further improvement, which we plan to explore in future work.

### C. Sensitive Analysis

*1) Impact of Various Occlusion Scenarios:* First, we evaluate model performance using point cloud data generated by the time-frequency ridge method [71] across four distinct scenarios. As shown in Table III, our method consistently achieves the highest F1-scores compared to other methods. Specifically, it demonstrates average improvements of 13.71% in the absence of obstacles, 8.10% with a clothes rack as the obstacle, 12.50% with a poster board, and 6.80% with a potted plant. These results highlight the robustness of our approach in handling environmental variability.

Second, we analyze the performance of our model using combined point cloud datasets from all four scenarios. As shown in Fig. 10, the average F1-scores are 0.97 for the empty scenario, 0.89 for the clothes rack scenario, 0.96 for the poster board

TABLE III
PID F1-SCORES ACROSS OCCLUSION SCENARIOS

| Method \ Scene | Empty | Clothes Rack | Poster Board | Pot Plant |
|---|---|---|---|---|
| mmGait [9] | 0.76 | 0.76 | 0.76 | 0.76 |
| MultiBranch [11] | 0.90 | 0.87 | 0.92 | 0.87 |
| MobileViT [37] | 0.90 | 0.84 | 0.88 | 0.87 |
| **Ours** | **0.97** | **0.89** | **0.96** | **0.89** |
| Improvement | 13.71% | 8.10% | 12.50% | 6.80% |

The best and second-best results highlighted in bold and underlined.

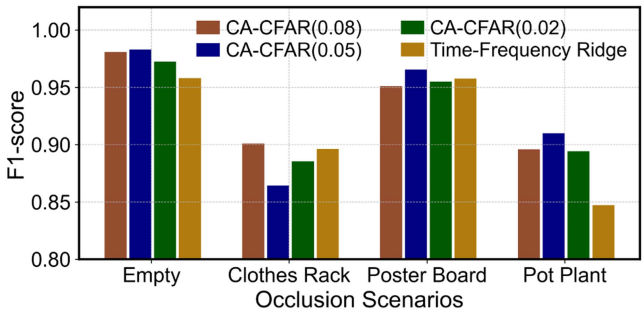



Fig. 10. Evaluation of the mmWave-PID F1-scores on four dataset scenarios (see Section VI-C). Colored bars represent different configurations: the CA-CFAR algorithm [70] at FARs of 0.02, 0.05, and 0.08, and the time-frequency ridge algorithm with a penalty factor of 0.05.

scenario, and 0.89 for the potted plant scenario. These findings further validate the robustness of our method across different environmental conditions.

Several conclusions can be drawn from Table III and Fig. 10: (1) The highest average F1-score of 0.96 occurs in the poster board scenario, where the obstacle completely blocks the subject. This demonstrates the effectiveness of our radar-based approach in complex situations where conventional cameras fail. (2) An average F1-score of 0.97 in the empty scenario suggests that the OSO-PID system can serve as a viable alternative to camera-based systems in typical environments. (3) Slightly lower F1-scores in the clothes rack and potted plant scenarios reflect the impact of material absorption and reflection properties on radar signals. Addressing these material-specific challenges will be a focus of our future research.

*2) Impact of Sample Augmentation:* To evaluate the effectiveness of our augmentation method, we conduct experiments on our mmWave radar-based PID method using several augmentation techniques. We compared our method with three alternatives: PatchAugment [44] and PointWolf [73], both of which augment point cloud samples through operations such as rotation and jittering, and Pointflow [61], which focuses on capturing motion information by calculating velocity differences. As shown in Table IV, all augmentation methods improved performance compared to using the original data across the four point cloud datasets. However, our proposed method demonstrated superior performance in enhancing results across diverse

TABLE IV
PID F1-SCORES FOR DIFFERENT POINT CLOUD AUGMENTATION METHODS

| Method \ Dataset | CFAR FAR:0.08 | CFAR FAR:0.05 | CFAR FAR:0.02 | Time-Frequency Ridge |
|---|---|---|---|---|
| PatchAugment [44] | 0.92 | 0.91 | 0.90 | 0.89 |
| PointWolf [73] | **0.93** | 0.91 | 0.91 | 0.91 |
| PointFlow [61] | 0.92 | 0.91 | 0.90 | 0.89 |
| Non-Augment | 0.92 | 0.89 | 0.89 | 0.89 |
| **Ours** | **0.93** | **0.93** | **0.93** | **0.92** |
| Improvement | 0.08% | 2.76% | 3.33% | 2.79% |

The best and second-best results highlighted in bold and underlined.

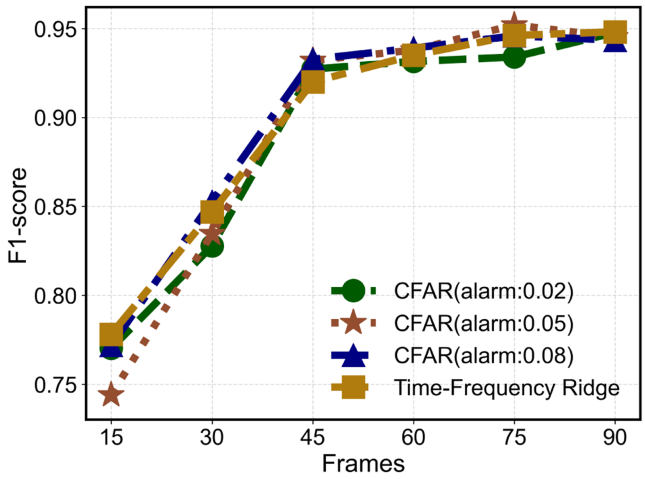



Fig. 11. F1-scores of mmWave-PID for different walking durations, showing improved performance with longer observation times.

radar point cloud datasets. Specifically, our method achieved average improvements of 0.08%, 2.76%, 3.33%, and 2.79%, respectively, outperforming the other augmentation approaches.

*3) Impact of Walking Time:* Extending radar sensing time during a person's walk intuitively increases the likelihood of accurate identification. We evaluated the F1-score for identifying 23 individuals using sequences of varying durations. As illustrated in Fig. 11, the F1-score improves significantly as the number of point cloud frames increases, ranging from 15 frames (F1-score: 0.77) to 45 frames (F1-score: 0.93), based on point cloud data generated with the CA-CFAR algorithm at a false alarm rate (FAR) of 0.08. However, the rate of improvement diminishes beyond 45 frames.

Balancing real-time prediction needs with high model performance, we conclude that using 45 point cloud frames per training or testing sample is optimal. This evaluation was extended to datasets generated using the CA-CFAR algorithm at FAR values of 0.02 and 0.05, as well as those derived from the time-frequency ridge method. In all cases, the results followed a similar trend, further validating this conclusion.

## D. Open-Set Recognition Results

In this section, the training dataset exclusively consists of a predetermined number of training samples for known identities,

TABLE V
F1-SCORES AT 9 OPENNESS LEVELS ON THE CA-CFAR DATASET [70] WITH FAR 0.05

| Method \ Openness | 6.19% | 8.01% | 9.73% | 11.36% | 12.90% | 14.36% | 15.76% | 17.08% | 18.35% |
|---|---|---|---|---|---|---|---|---|---|
| CIP [23] | 0.85 | 0.83 | 0.80 | 0.78 | 0.76 | 0.74 | 0.72 | 0.70 | 0.68 |
| OpenMax [21] | 0.85 | 0.83 | 0.80 | 0.78 | 0.76 | 0.74 | 0.72 | 0.70 | 0.68 |
| Weibull-based | 0.86 | 0.84 | 0.81 | 0.78 | 0.76 | 0.74 | 0.72 | 0.70 | 0.68 |
| Softmax-based | 0.88 | 0.85 | 0.82 | 0.80 | 0.77 | 0.75 | 0.73 | 0.71 | 0.69 |
| **Ours** | **0.89** | **0.87** | **0.86** | **0.84** | **0.83** | **0.81** | **0.79** | **0.79** | **0.77** |

We mark the best and second-best results using bold and underlined text, respectively.

TABLE VI
F1-SCORES AT 9 OPENNESS LEVELS ON THE MMGAIT DATASET [9]

| Method \ Openness | 6.19% | 8.01% | 9.73% | 11.36% | 12.90% | 14.36% | 15.76% | 17.08% | 18.35% |
|---|---|---|---|---|---|---|---|---|---|
| CIP [23] | 0.80 | 0.78 | 0.76 | 0.74 | 0.73 | 0.71 | 0.71 | 0.70 | 0.69 |
| OpenMax [21] | 0.81 | 0.79 | 0.73 | 0.71 | 0.69 | 0.67 | 0.66 | 0.64 | 0.63 |
| **Ours** | **0.82** | **0.80** | **0.78** | **0.76** | **0.74** | **0.73** | **0.72** | **0.71** | **0.71** |

We mark the best and second-best results using bold and underlined text, respectively.

whereas the test dataset contains test data for all subjects (known and unknown). Our experiments are conducted using a range of openness configurations, aiming to mimic real-world scenarios.

*1) Overall Performance:* In our experiments, we compare our method with two representative open-set methods: CIP [23] and OpenMax [21]. The first method leverages Dopper image obtained from the mmWave radar for the open-set PID task. The second method is widely regarded as a backbone for open-set tasks. Note that, due to CIP and OpenMax being designed for images, we substitute the feature extractor with our PID network (without the contrastive learning module), keeping judgment rules unchanged. Furthermore, we perform a comprehensive evaluation of our method through ablation experiments, incorporating judgments based solely on the Weibull model or the softmax approach, respectively. The results are shown in Table V.

According to Table V, the comparison results yield several notable observations. First, we observe a decrease in the performance of all algorithms with the level of openness increasing. This decline is expected since the identification task becomes more challenging with a larger number of unknown identities. However, our method consistently achieves the highest performance across all levels of openness and gracefully decreases in comparison to the other two methods. Second, we observe the efficacy of supervised contrastive learning for the open-set task through a comparative analysis with OpenMax [21]. The performance gap widens as the level of openness increases, reaching a maximum difference of 0.09 at an openness level of 18.35%. This indicates that the contrastive learning method effectively enhances performance for open-set tasks. Third, we observe that both softmax-based and Weibull model-based methods are beneficial for the open-set task, especially when the level of openness is low. For example, in our experiment, we define that if the maximum softmax probability is less than 0.4, the class label is assigned as "unknown". As shown in Table V, when the level of openness is below 11.36%, the softmax-based method achieves a performance exceeding 0.8. Hence, we incorporate two constraints in our method to identify the unknown class. The first constraint is that the maximum posterior probability is less than a hyper-parameter denoted as $\delta$, set to 0.4 in our experiment. The second constraint is derived from the index corresponding to the maximum posterior probability from the Weibull model, which is set to 12, indicating the label for unknown classes.

*2) Performance on the mmGait Dataset:* To verify the ability of our method, we further perform our method on the public mmGait dataset [9]. As shown in Table VI, our method demonstrates superior performance compared to the two methods. Furthermore, we observe that while the OpenMax [21] method achieves similar performance to ours at low openness levels, our method exhibits a more gradual decline in performance as the openness increases. Concurrently, even though the CIP method demonstrates a gradual decline in performance, our method consistently outperforms it at every level of openness.

*3) Robustness Performance on Two Datasets:* We further apply our method to various radar point cloud datasets, which include CA-CFAR-based (FAR: 0.02 and 0.08) and the time-frequency ridge-based dataset. Based on the results presented in Fig. 12, it is evident that our method consistently achieves superior performance compared to the other two methods on all datasets. Moreover, the results yield several notable observations. First, we observe that the performance difference between our method and the compared methods increases with the level of openness on all datasets. This indicates that our method more effectively addresses the open-set task compared to other methods, especially when the level of openness is high. Second, as shown in Table I, we establish 9 distinct openness levels. We observe that the performance difference of our method between the minimum and maximum openness levels achieves

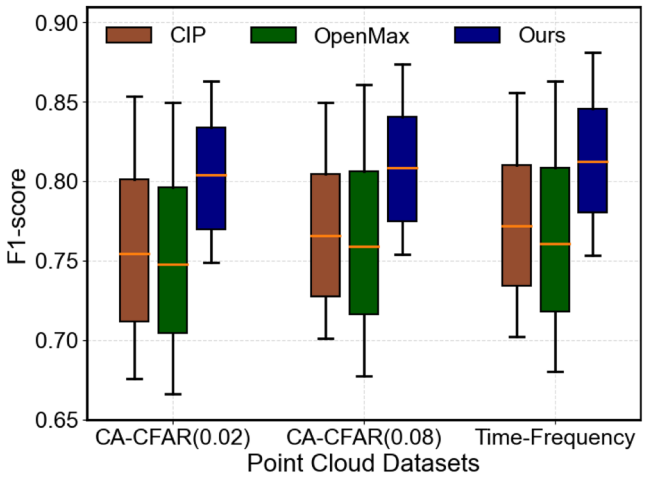
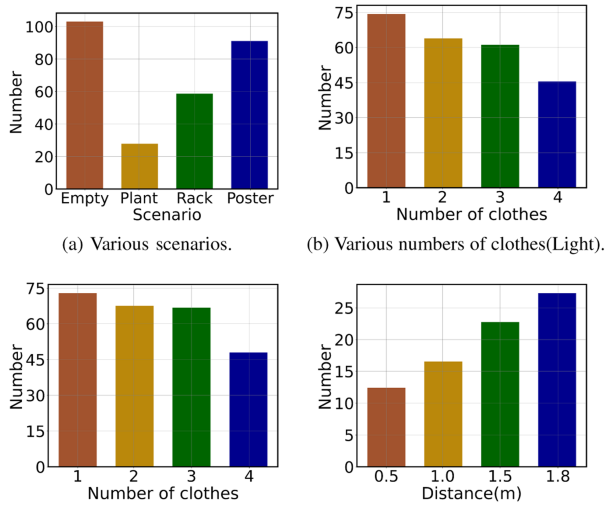



Fig. 12. Performance on the open-set PID task across 9 levels of openness and three datasets. Performance declines as openness increases, with our method consistently outperforming others.




(a) Various scenarios. (b) Various numbers of clothes(Light). (c) Various numbers of clothes(Thick). (d) Various distances to the radar.

Fig. 13. Average number of points in each point cloud: (a) calculated from point clouds obtained under various occlusion scenarios; (b) from point clouds using a clothes rack with varying numbers of light clothing (e.g., shirts) as the obstacle; (c) from point clouds using a clothes rack with varying numbers of thick clothing (e.g., coats) as the obstacle; and (d) from point clouds using a potted plant at different distances from the radar as the obstacle.

0.11, 0.12 and 0.13, respectively, on the three radar point cloud datasets. These results demonstrate that the performance of our method decreases by an average of 1.33% as the degree of openness increases by approximately 1.35%. This degradation is less compared to other methods, highlighting the robustness of our approach across various point cloud datasets and different openness levels.

*4) Performance on Different Distributions:* According to EVT, the distribution of extreme values converges to one of three forms: the Gumbel, Frechet, or Weibull distribution. To validate the efficacy of the Weibull distribution, we conducted evaluations on the mmWave-ocPID and mmGait [9] datasets, comparing the performance of different EVT distributions.

The results show that the Weibull distribution outperforms the Gumbel and Frechet distributions on both datasets. For instance, for the mmWave-ocPID dataset, the Weibull distribution achieved an average F1-score of 0.83 across 9 openness levels, while the Gumbel and Frechet distributions have scores of only 0.14. Similarly, using the mmGait dataset, the Weibull distribution achieved an average F1-score of 0.75, significantly surpassing the score of 0.12 from the Gumbel and Frechet distributions. Our evaluation results are consistent with theoretical works in [66] and [67], both showing the efficacy of the Weibull distribution for open-set recognition.

### *E. Discussion and Future Work*

By using our mmWave-ocPID dataset, we compare the average number of points in each radar point cloud to analyze mmWave signal attenuation across different occlusion scenarios. Moreover, we use a clothes rack and a potted plant as obstacles, respectively. Clothes with different materials and thicknesses are hung on the rack, and the potted plant is placed at different distances from the radar. A volunteer is instructed to walk behind the obstacle for 4 consecutive minutes to collect mmWave signals. These signals are further processed by the fast fourier transform (FFT) [60] and cell-averaging constant false alarm rate (CA-CFAR) [70] algorithms to generate radar point clouds for analysis, where the false alarm rate in CA-CFAR is set to 0.05. With the data from our new experiments, we analyze the effects of the number of obstacles, e.g., clothes, as well as the obstacle distance from the radar on the number of mmWave radar point cloud data. The results are shown in Fig. 13. As shown in Fig. 13(a), the average number of points in each point cloud varies with different occlusion scenarios, confirming that radar signals experience different levels of attenuation caused by different obstacles. As shown in Fig. 13(b) and (c), the average number of points in each radar point cloud decreases as the number of clothes increases, verifying that a larger obstacle dimension results in greater signal attenuation. Furthermore, as shown in Fig. 13(d), as the distance to the radar decreases, mmWave signals experience greater attenuation, reducing the number of points in the corresponding point clouds.

In addition, we plan to further improve our method and our future works include the following: (1) The assembly of an expansive mmWave radar dataset, featuring a diverse array of subjects exhibiting various gaits and motion patterns. This comprehensive dataset will facilitate an in-depth exploration. (2) The fusion of mmWave radar with vision-based algorithms [74], to pave the way toward an extended multi-modality person identification framework. (3) The incorporation of incremental learning and unsupervised learning technologies [75] to accentuate the differentiation of unknown classes in the unsupervised open-set problem.

## VII. Conclusion

In this paper, we investigate the feasibility of employing mmWave radar for PID in heavily or completely occluded scenarios, while exploring the practical open-set PID problem. We first build a mmWave radar dataset comprising four occlusion

scenarios with 23 participants. We then propose a novel person identification approach incorporating a data augmentation strategy and a supervised contrastive learning method. The supervised contrastive learning not only enhances the robustness of the approach but also constructs a compact feature space for the subsequent open-set PID task. Our approach is specifically designed to handle heavy or even full occlusions that can well compensate for vision-based PID methods in complex environments. For the open-set recognition problem, our approach integrates statistical models with a pre-trained model based on super contrastive learning. Extensive experiments show that our method achieves 0.93, 0.93, 0.93, and 0.92 PID F1-scores at various occlusion scenarios, which are superior to the state-of-the-art methods.

## REFERENCES


[1] J. Pegoraro, J. O. Lacruz, F. Meneghello, E. Bashirov, M. Rossi, and J. Widmer, "RAPID: Retrofitting IEEE 802.11 ay access points for indoor human detection and sensing," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4501–4519, May 2024.

[2] D. Salami, R. Hasibi, S. Palipana, P. Popovski, T. Michoel, and S. Sigg, "Tesla-rapture: A lightweight gesture recognition system from mmWave radar sparse point clouds," *IEEE Trans. Mobile Comput.*, vol. 22, no. 8, pp. 4946–4960, Aug. 2023.

[3] L. Deng, J. Yang, S. Yuan, H. Zou, C. X. Lu, and L. Xie, "GaitFi: Robust device-free human identification via WiFi and vision multimodal learning," *IEEE Internet Things J.*, vol. 10, no. 1, pp. 625–636, Jan. 2023.

[4] J. Zhang et al., "A survey of mmWave-based human sensing: Technology, platforms and applications," *IEEE Commun. Surveys Tut.*, vol. 25, no. 4, pp. 2052–2087, Fourth Quarter 2023.

[5] A. Sepas-Moghaddam, S. Ghorbani, N. F. Troje, and A. Etemad, "Gait recognition using multi-scale partial representation transformation with capsules," in *Proc. IEEE 25th Int. Conf. Pattern Recognit.*, 2020, pp. 8045–8052.

[6] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Geinet: View-invariant gait recognition using a convolutional neural network," in *PRoc. IEEE Int. Conf. Biometrics*, 2016, pp. 1–8.

[7] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, "MU-ID: Multi-user identification through gaits using millimeter wave radios," in *Proc. IEEE Conf. Comput. Commun.*, 2020, pp. 2589–2598.

[8] P. Zhao et al., "mID: Tracking and identifying people with millimeter wave radar," in *Proc. IEEE 15th Int. Conf. Distrib. Comput. Sensor Syst.*, 2019, pp. 33–40.

[9] Z. Meng et al., "Gait recognition for co-existing multiple people using millimeter wave sensing," in *Proc. Conf. Assoc. Advance. Artif. Intell.*, 2020, pp. 849–856.

[10] Y. Cheng and Y. Liu, "Person reidentification based on automotive radar point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5101913.

[11] Z. Xia, G. Ding, H. Wang, and F. Xu, "Person identification with millimeter-wave radar in realistic smart home scenarios," *IEEE Geosci. Remote. Sens. Lett.*, vol. 19, 2022, Art. no. 3509405.

[12] V. B. Semwal, A. Mazumdar, A. Jha, N. Gaud, and V. Bijalwan, "Speed, cloth and pose invariant gait recognition-based person identification," in *Machine Learning: Theoretical Foundations and Practical Applications*. Berlin, Germany: Springer, 2021, pp. 39–56.

[13] Z. Sun, Z. Yu, Q. Wang, Z. Wang, B. Guo, and H. Zhang, "CovertEye: Gait-based human identification under weakly constrained trajectory," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5558–5570, May 2024.

[14] X. Zhang, D. Zhang, Y. Xie, D. Wu, Y. Li, and D. Zhang, "Waffle: A waterproof mmWave-based human sensing system inside bathrooms with running water," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 4, pp. 1–29, 2024.

[15] R. Feng, E. De Greef, M. Rykunov, H. Sahli, S. Pollin, and A. Bourdoux, "Multipath ghost recognition for indoor MIMO radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5104610.

[16] E. Kobayashi, A. Kosuge, M. Hamada, and T. Kuroda, "An occlusion-resilient mmWave imaging radar-based object recognition system using synthetic training data generation technique," in *Proc. IEEE 49th Annu. Conf. Ind. Electron. Soc.*, 2023, pp. 1–6.

[17] J. Yang, K. Zhou, Y. Li, and Z. Liu, "Generalized out-of-distribution detection: A survey," *Int. J. Comput. Vis.*, vol. 132, pp. 5635–5662, 2024.

[18] S. Vaze, K. Han, A. Vedaldi, and A. Zisserman, "Open-set recognition: A good closed-set classifier is all you need," in *Proc. 10th Int. Conf. Learn. Representations*, 2022, pp. 1–26. [Online]. Available: https://openreview.net/pdf?id=5hLP5JY9S2d

[19] C. Geng, S.-J. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3614–3631, Oct. 2021.

[20] A. U. Rehman, W. Jiao, J. Sun, H. Pan, and T. Yan, "Open set recognition methods for fault diagnosis: A review," in *Proc. IEEE 15th Int. Conf. Adv. Comput. Intell.*, 2023, pp. 1–8.

[21] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1563–1572.

[22] S. Kong and D. Ramanan, "OpenGAN: Open-set recognition via open data generation," in *Proc. 2021 IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 793–802.

[23] Z. Ni and B. Huang, "Open-set human identification based on gait radar micro-doppler signatures," *IEEE Sensors J.*, vol. 21, no. 6, pp. 8226–8233, Mar. 2021.

[24] P. Khosla et al., "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 18 661–18 673.

[25] A. Vaswani et al., "Attention is all you need," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[27] K. He, H. Fan, Y. Wu, S. Xie, and R. B. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. 2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9726–9735.

[28] L. Y. Wu et al., "Learning resolution-adaptive representations for cross-resolution person re-identification," *IEEE Trans. Image Process.*, vol. 32, pp. 4800–4811, 2023.

[29] D. Liu et al., "Generative metric learning for adversarially robust open-world person re-identification," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 19, no. 1, pp. 1–19, 2023.

[30] P. Cao, W. Xia, M. Ye, J. Zhang, and J. Zhou, "Radar-ID: Human identification based on radar micro-doppler signatures using deep convolutional neural networks," *IET Radar, Sonar Navigation*, vol. 12, no. 7, pp. 729–734, 2018.

[31] B. Vandersmissen et al., "Indoor person identification using a low-power FMCW radar," *IEEE Trans. Geosci. Remote. Sens.*, vol. 56, no. 7, pp. 3941–3952, Jul. 2018.

[32] Y. Yang, C. Hou, Y. Lang, G. Yue, Y. He, and W. Xiang, "Person identification using micro-doppler signatures of human motions and UWB radar," *IEEE Microw. Wirel. Compon. Lett.*, vol. 29, no. 5, pp. 366–368, May 2019.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[34] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. 2017 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.

[35] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021, pp. 1–21. [Online]. Available: https://openreview.net/pdf?id=YicbFdNTTy

[36] B. Heo, S. Yun, D. Han, S. Chun, J. Choe, and S. J. Oh, "Rethinking spatial dimensions of vision transformers," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 11 916–11 925.

[37] S. Mehta and M. Rastegari, "MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer," in *Proc. Int. Conf. Learn. Representations*, 2022, pp. 1–26. [Online]. Available: https://openreview.net/pdf?id=vh-0sUt8HlG

[38] Z. Zhang, H. Zhang, L. Zhao, T. Chen, S. Ö. Arik, and T. Pfister, "Nested hierarchical transformer: Towards accurate, data-efficient and interpretable visual understanding," in *Proc. Conf. Assoc. Advance. Artif. Intell.*, 2022, pp. 3417–3425.

[39] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 77–85.

[40] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5099–5108.

[41] Y. Chen et al., "PointMixup: Augmentation for point clouds," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 330–345.

[42] V. Verma et al., "Manifold mixup: Better representations by interpolating hidden states," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2019, pp. 6438–6447.
[43] R. Li, X. Li, P. Heng, and C. Fu, "PointAugment: An auto-augmentation framework for point cloud classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6377–6386.
[44] S. V. Sheshappanavar, V. V. Singh, and C. Kambhamettu, "Patchaugment: Local neighborhood augmentation in point cloud classification," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2021, pp. 2118–2127.
[45] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1757–1772, Jul. 2013.
[46] L. Neal, M. L. Olson, X. Z. Fern, W. Wong, and F. Li, "Open set learning with counterfactual images," in *Proc. 15th Eur. Conf. Comput. Vis.*, Springer, 2018, pp. 620–635.
[47] R. Yoshihashi, W. Shao, R. Kawakami, S. You, M. Iida, and T. Naemura, "Classification-reconstruction learning for open-set recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4016–4025.
[48] P. Oza and V. M. Patel, "C2AE: Class conditioned auto-encoder for open-set recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2307–2316.
[49] X. Sun, Z. Yang, C. Zhang, K. V. Ling, and G. Peng, "Conditional gaussian distribution learning for open set recognition," in *Proc. 2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 13 477–13 486.
[50] Y. Shu, Y. Shi, Y. Wang, T. Huang, and Y. Tian, "P-ODN: Prototype-based open deep network for open set recognition," *Sci. Rep.*, vol. 10, no. 1, 2020, Art. no. 7146.
[51] G. Chen et al., "Learning open set network with discriminative reciprocal points," in *Proc. 16th Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 507–522.
[52] Y. Yang, Y. Ge, B. Li, Q. Wang, Y. Lang, and K. Li, "Multiscenario open-set gait recognition based on radar micro-doppler signatures," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 2519813.
[53] C. Liu, S. Liu, C. Zhang, Y. Huang, and H. Wang, "Multipath propagation analysis and ghost target removal for FMCW automotive radars," in *Proc. IET Int. Radar Conf.*, 2020, pp. 330–334.
[54] H. A. Obeidat, A. A. Abdullah, M. F. Mosleh, A. Ullah, O. A. Obeidat, and R. A. Abd-Alhameed, "Comparative study on indoor path loss models at 28 ghz, 60 ghz, and 73.5 ghz frequency bands," *Appl. Comput. Electromagn. Soc. J.*, vol. 35, pp. 119–128, 2020.
[55] M. Khatun, C. Guo, and H. Mehrpouyan, "Penetration and reflection characteristics in millimeter-wave indoor channels," in *Proc. 2021 IEEE-APS Topical Conf. Antennas Propag. Wirel. Commun.*, 2021, pp. 1–5.
[56] N. Scheiner et al., "Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using doppler radar," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2068–2077.
[57] Y. Xu, G. Liu, and T. Jiang, "Leveraging rough-relay-surface scattering for non-line-of-sight mmWave radar sensing," *IEEE Internet Things J.*, vol. 11, no. 6, pp. 10964–10978, Mar. 2024.
[58] J. Tian, Q. Zhang, and F. Yu, "Non-coherent detection for two-way AF cooperative communications in fast Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2753–2762, Oct. 2011.
[59] S. Naskar and A. K. Dutta, "RadRCom: A relay-assisted radar communication system design framework," *IEEE Access*, vol. 12, pp. 72635–72649, 2024.
[60] X. Li, X. Wang, Q. Yang, and S. Fu, "Signal processing for TDM MIMO FMCW millimeter-wave radar sensors," *IEEE Access*, vol. 9, pp. 167 959–167 971, 2021.
[61] Y. Huang et al., "HDNet: Hierarchical dynamic network for gait recognition using millimeter-wave radar," 2022, *arXiv:2211.00312*.
[62] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.
[63] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2317–2324, Nov. 2014.
[64] S. Kotz and S. Nadarajah, *Extreme Value Distributions: Theory and Applications*. Singapore: World Scientific, 2000.
[65] W. J. Scheirer, A. Rocha, R. J. Micheals, and T. E. Boult, "Meta-recognition: The theory and practice of recognition score analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1689–1695, Aug. 2011.
[66] E. M. Rudd, L. P. Jain, W. J. Scheirer, and T. E. Boult, "The extreme value machine," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 762–768, Mar. 2018.
[67] P. Zhang, J. Wang, A. Farhadi, M. Hebert, and D. Parikh, "Predicting failures of vision systems," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3566–3573.
[68] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. London, U.K.: Springer, 2001.
[69] W. J. Scheirer, A. Rocha, R. J. Micheals, and T. E. Boult, "Meta-recognition: The theory and practice of recognition score analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1689–1695, Aug. 2011.
[70] H. M. Finn and R. S. Johhnson, "Adaptive detection mode with threshold control as a function of spatially sampled clutter level estimates," *RCA Rev.*, vol. 29, no. 3, pp. 414–464, Sep. 1968.
[71] D. Iatsenko, P. V. E. McClintock, and A. Stefanovska, "Extraction of instantaneous frequencies from ridges in time-frequency representations of signals," *Signal Process.*, vol. 125, pp. 290–303, 2016.
[72] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
[73] S. Kim, S. Lee, D. Hwang, J. Lee, S. J. Hwang, and H. J. Kim, "Point cloud augmentation with weighted local transformations," in *Proc. Int. Conf. Comput. Vis.*, 2021, pp. 528–537.
[74] D. Liu, L. Wu, F. Zheng, L. Liu, and M. Wang, "Verbal-person nets: Pose-guided multi-granularity language-to-person generation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8589–8601, Nov. 2023.
[75] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. Van De Weijer, "Class-incremental learning: Survey and performance evaluation on image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5513–5533, May 2023.

**Tao Wang** (Student Member, IEEE) received the BS degree in computer science and technology from Jilin University, in 2019 and the MS degree in computer science and technology from Harbin Institute of Technology, respectively, in 2021. He is currently working toward the PhD degree with International Research Institute for Artificial Intelligence, Harbin Institute of Technology, Shenzhen. His current research interests include wireless sensing and artificial intelligence.

**Yang Zhao** (Senior Member, IEEE) received the BS degree in electrical engineering from Shandong University, in 2003, the MS degree in electrical engineering from the Beijing University of Aeronautics and Astronautics, in 2006, and the PhD degree in electrical and computer engineering from the University of Utah, in 2012. He was a lead research engineer at GE Global Research between 2013 and 2021. Since 2021, he has been with the Harbin Institute of Technology, Shenzhen, where he is a research professor with the International Research Institute for Artificial Intelligence. His research interests include wireless sensing, edge computing and cyber physical systems.

**Ming-Ching Chang** (Senior Member, IEEE) received the BS degree in civil engineering and the MS degree in computer science and information engineering (CSIE) from the National Taiwan University in 1998 and 1996, respectively, and the PhD degree from the Laboratory for Engineering Man/Machine Systems (LEMS), School of Engineering, Brown University, in 2008. From 2008 to 2016, he was a computer scientist with the GE Global Research Center. He is currently an associate professor with the Department of Computer Science, College of Engineering and Applied Sciences (CEAS), University at Albany, State University of New York (SUNY). His research interests include video analytics, computer vision, image processing, and artificial intelligence.

**Jie Liu** (Fellow, IEEE) is a chair professor with the Harbin Institute of Technology Shenzhen (HIT Shenzhen), China and the dean of its AI Research Institute. Before joining HIT, he spent 18 years with Xerox PARC and Microsoft. He was a principal research manager with Microsoft Research, Redmond and a partner of the company. His research interests are Cyber-Physical Systems, AI for IoT, and energy-efficient computing. He received IEEE TC-CPS Distinguished Leadership Award and six Best Paper Awards from top conferences. He is an ACM distinguished scientist, and founding chair of ACM SIGBED China.