Unsupervised Domain Adaptation for Disguised-gait-based Person Identification on Micro-Doppler Signatures

Yang, Yang ; Yang, Xiaoyi ; Sakamoto, Takuya ; Fioranelli, Francesco; Li, Beichen ; Lang, Yue

DOI
10.1109/TCSVT.2022.3161515

Publication date
2022

Document Version
Final published version

Published in
IEEE Transactions on Circuits and Systems for Video Technology

Unsupervised Domain Adaptation for Disguised-Gait-Based Person Identification on Micro-Doppler Signatures

Yang Yang, Xiaoyi Yang, Takuya Sakamoto, *Senior Member, IEEE*, Francesco Fioranelli, *Senior Member, IEEE*, Beichen Li, and Yue Lang

***Abstract*— In recent years, gait-based person identification has gained significant interest for a variety of applications, including security systems and public security forensics. Meanwhile, this task is faced with the challenge of disguised gaits. When a human subject changes what he or she is wearing or carrying, it becomes challenging to reliably identify the subject's identity using gait data. In this paper, we propose an unsupervised domain adaptation (UDA) model, named *Guided Subspace Alignment under the Class-aware condition (G-SAC)*, to recognize human subjects based on their disguised gait data by fully exploiting the intrinsic information in gait biometrics. To accomplish this, we employ neighbourhood component analysis (NCA) to create an intrinsic feature subspace from which we can obtain similarities between normal and disguised gaits. With the aid of a proposed constraint for adaptive class-aware alignment, the class-level discriminative feature representation can be learned guided by this subspace. Our experimental results on a measured micro-Doppler radar dataset demonstrate the effectiveness of our approach. The comparison results with several state-of-the-art methods indicate that our work provides a promising domain adaptation solution for the concerned problem, even in cases where the disguised pattern differs significantly from the normal gaits. Additionally, we extend our approach to more complex multi-target domain adaptation (MTDA) challenge and video-based gait recognition tasks, the superior results demonstrate that the proposed model has a great deal of potential for tackling increasingly difficult problems.**

***Index Terms*— Micro-Doppler signatures, gait recognition, radar-based person identification, transfer learning, unsupervised domain adaptation.**

Manuscript received 18 December 2021; revised 26 February 2022; accepted 17 March 2022. Date of publication 22 March 2022; date of current version 6 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62101378 and Grant 62171318, in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 19H02155 and Grant 21H03427, in part by the Japan Science and Technology Agency (JST) PRESTO under Grant JPMJPR1873, in part by the JST Center of Innovation (COI) under Grant JPMJCE1307, and in part by the SECOM Science and Technology Foundation. This article was recommended by Associate Editor Y. M. Ro. *(Corresponding author: Beichen Li.)*

Yang Yang, Xiaoyi Yang, and Beichen Li are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: wednesday33@126.com).

Takuya Sakamoto is with the Graduate School of Engineering, Kyoto University, Kyoto 615-8510, Japan.

Francesco Fioranelli is with the Department of Microelectronics, Delft University of Technology, 2628 CD Delft, The Netherlands.

Yue Lang is with the School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSVT.2022.3161515.

Digital Object Identifier 10.1109/TCSVT.2022.3161515

I. Introduction

PERSON identification has increasingly become a hot research topic in numerous applications. Due to its unique advantages that there is no need for cooperation from human subjects, gait biometric is more suitable for the person identification task than the common biometrics such as iris [1], fingerprint [2], and face [3]. In [4], the authors reveal the uniqueness of individual gait patterns in clinical biomechanics. The most widely used sensors for biometric-based person identification include cameras [5], LiDAR [6], and wearable sensors [7]. Nevertheless, vision-based and LiDAR-based identification systems will heavily suffer from the environments (such as low light, bad weather, occlusion, etc.) and cameras easily cause the risk of privacy leakage. Wearable sensors cannot be used for non-cooperative targets. In recent years, people have investigated the feasibility of gait recognition based on Wi-Fi signals because of the advantages of low-cost, non-intrusiveness, and light-insensitivity [8], [9]. On the other hand, due to channel congestion, Wi-Fi signals are highly susceptible to interference and environmental noise [10]. Thus, it is desirable to investigate a sensor capable of detecting human movement in complex settings.

Thanks to the micro-Doppler phenomenon [11], radar has proved to be effective in gait-based person identification under complex environments [12]. Micro-Doppler, also commonly dubbed as micro-Doppler signatures (m-Ds), is the additional frequency modulation of raw radar echoes, and this modulation is caused by the micro-motions of the human body parts (e.g., rotation of the torso, the periodic swing of the arms and legs) when the subject moves in the radial direction of radar illumination [11]. Abdulatif *et al.* [13] investigate how the m-Ds are affected by the body mass index (BMI), which is a key factor that causes the uniqueness of different human gaits, and successfully recognize 22 human subjects with an overall 98% accuracy on the unseen test set.

In recent years, several algorithms have been proposed to use m-Ds for gait-based person identification [14]–[18]. However, these studies do not account for disguised states (e.g., clothing, different moving speeds, whether carrying objects, etc.). In this case, the person identification model will easily suffer from the numerous unknown disguised human gaits because the radar echo intensities, Doppler bandwidth or period of motion will be altered when a test subject changes his or her moving states. Basically, changing to a

new moving state means that the original distribution of human gait data will be shifted as well [19]. For example, the authors of [19] investigate recognizing the subjects in different indoor environments, and the discrepancies between the normal and disguised gaits most likely lead to misclassification. Therefore, the m-Ds in human backscattering echoes are varied by the appearance of the subject, and it is quite challenging to identify the subject when its m-Ds change.

The majority of approaches fail on this challenge because they are based on the assumption that the data for normal gaits and disguised gaits follow the independently identical distribution (i.i.d) hypothesis. For the non-i.i.d problem, if the gap between the new situations (a.k.a. target domain) and the original scenarios (a.k.a. source domain) is not taken into account, the classification model's performance will be significantly deteriorated.

This challenge is referred to as a 'cross-domain' problem, and this problem enables a fresh research field of domain adaptation (DA) [20]. The objective of DA is to bridge the domain gap between the source and target domains so that the knowledge in the source domain can be applied to the task in the target domain. DA issues can be addressed in a supervised or unsupervised manner, the former requires adequate labeled data, whereas the latter does not. Annotating images is both time-consuming and labor-intensive in practical scenarios, researchers are likely to undertake domain adaptation unsupervised rather than supervised [21]. If we train an identification model in a supervised way, we must label a large number of samples in new scenarios, which becomes more difficult when the number of new scenarios is enormous.

A large number of UDA approaches make great progress in typical computer vision tasks such as image classification [22], [23], person re-identification (Re-ID) [21], [24], [25], and semantic segmentation [26]. To be specific, researchers employ the constraints of adversarial training [27], [28], statistical distance [29], and clustering methods [21], [30] to ensure the projected features of source and target domain data can be well aligned. Hence, UDA is considered a feasible solution for person identification based on disguised gaits.

During the experimental exploration, we observed that the discrepancy between the normal domain and different target domains is not treated equally. This is because disguised gaits differ in multiple scenarios, the discrepancy between various disguised gaits and normal gaits will also differ. As depicted in Fig. 1, some disguised scenarios deviated less from the normal scenario, and some scenarios differed considerably from the normal scenario. This observation inspires us that such differences in the discrepancy should be considered during domain adaptation.

In this paper, we propose a new UDA model for person identification based on the disguised gaits using the radar sensor. In our model, the original features of the source and target images are learned under supervised and unsupervised constraints. Considering the differences in the discrepancy between normal and various disguised gaits, we use the neighbourhood components analysis (NCA) [31] principle to guide the learning of the *intrinsic* feature subspace. The primary design of our model is shown in Fig. 2. As shown in this figure,

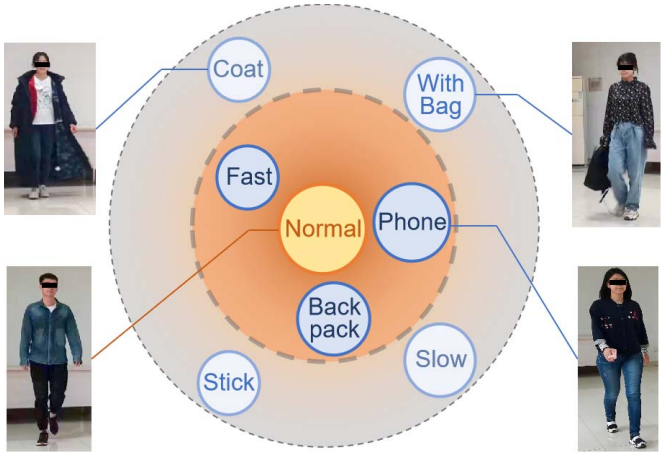


Fig. 1. Illustration of the differences in the discrepancy between disguised gaits and normal gaits. The orange circle in the center represents the normal gaits, and the other blue circles indicate the disguised gaits. The location of each disguised gait, i.e., in the inner or outer ring, represents the similarity between the normal gaits and the corresponding disguised gaits.

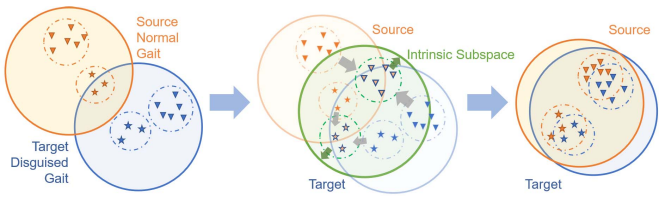


Fig. 2. Illustration of the proposed approach. The orange circle (solid line) is the source domain (feature subspace), and the blue circle (solid line) indicates the target domain (feature subspace). The triangles and stars in the circles indicate the different classes in each domain, and the dash-dotted line circles the instances of the same class. The green circle is the proposed NCA-guided intrinsic subspace. The gray and green arrows indicate the intra-class and inter-class alignment processes, and the size of the arrows represents the self-adaptive alignment weights.

the guided intrinsic subspace integrates both the intra-class similarity information and the inter-class discrepancy information between the source and the target domains, as the features of different domains are well-aligned at the class level. In order to maintain the precise cross-domain alignment in the guided subspace, we design a new self-adaptive class-aware alignment constraint. We name our model as *Guided Subspace Alignment under the Class-aware condition (G-SAC)*.

Our contributions are summarized as follows:

- To the best of our knowledge, this work is the first exploration that addresses the issue of person identification based on disguised gaits using radar m-Ds. All the previous research reported in the literature does not consider the case of the disguised gaits.
- We propose a novel neural-network-based model 'G-SAC' to solve the problem. To overcome the distribution shifts at the class level in adaptation, G-SAC employs supervised and unsupervised constraints to maintain the consistency of the class-aware data distributions of both source and target data in the original subspace.
- We notice the differences in the discrepancy between normal and various disguised gaits and propose an NCA-guided intrinsic subspace learning scheme to avoid

the risk of inappropriate transfer caused by this discrepancy. In addition, a self-adaptive class-aware alignment constraint is designed for the optimization so that the training process would not be impacted by the imbalance within each class and between different classes.

- We build a monostatic radar system and establish a dataset containing both normal and disguised gaits captured by this system. We also evaluate the G-SAC model on this dataset, then use the state-of-the-art (SOTA) UDA methods in computer vision tasks for comparison. The experimental results show that the G-SAC outperforms the other methods and is effective for disguised-gait-based person identification.

The rest of the paper is structured as follows. Section II briefly reviews related work. Section III presents the radar system and the measurement of radar gaits in different scenarios. In Section IV, we describe the proposed G-SAC algorithm in detail. Section V presents the experiments and discussion. Finally, we conclude the paper in Section VI.

II. Related Work

A. Radar-Based Person Identification

In recent years, many researchers have explored a number of radar-based person identification methods. Trommel *et al.* [16] used a radar system to detect human gaits and employed a deep convolutional neural network that is able to recognize whether the human target is detected and to identify the number of detected human targets. In [14], a neural-network-based vote algorithm was presented for gait-based person identification using m-Ds. Chen *et al.* [15] utilized m-Ds to distinguish armed and unarmed people based on a multi-static radar system. In [17], a 25 GHz frequency modulated continuous wave (FMCW) single-input-single-output (SISO) radar was employed in industrial safety for real-time human-robot identification. Papanastasiou *et al.* [18] recognized human individuals based on radar measurements using an X-band radar. In [32], a low-power FMCW radar system was established to automatically identify spontaneous walkers in different rooms. In [13], the convolutional autoencoder was used to extract the latent representations from the m-Ds for person identification. Lang *et al.* [33] used a mono-static radar system for joint motion classification and person identification. The reservoir computing networks were explored to identify a limited number of people in an indoor environment [34].

These researches seek to build an effective system for person identification based on gait, or to develop the capability of using various radars or human movements for person identification, but, to the best of our knowledge, none of them considered the problem of disguised gaits in the classification process.

B. Domain Adaptation

In terms of whether the labels of the data in the target domain are accessible, domain adaptation can be divided into supervised DA, semi-supervised DA, and unsupervised DA [20]. The UDA is the most challenging task among them, as the target labels cannot be accessed at all.

1) Alignment-Based UDA: Many studies aimed to learn domain-invariant representations by establishing the objective function with different metric criteria, e.g., maximum mean discrepancy (MMD) [35], CORAL [29], Kullback-Leibler (KL) divergence [36], and Wasserstein distance [37]. Zhu *et al.* [23] proposed a new local distribution metric to measure the discrepancy between relevant subdomains within the same category of different domains. Taking advantage of the adversarial learning strategy of the GAN, GAN-based UDA methods [27], [38] have attempted to make a domain invariant feature space using the well-trained generator. Luo *et al.* [26] designed a GAN model to alleviate the negative transfer during the feature alignment. Cui *et al.* [28] proposed a new network to reduce domain-specific representations. In [22], the nuclear-norm was considered as an objective function to enhance the discriminability and diversity of the features. Xu *et al.* [39] gradually encouraged feature-norm enlargement across domains for each sample.

2) Clustering-Based UDA: A variety of methods [30], [39] show that maintaining the structure of the target domain is necessary during domain adaptation, and clustering is a widely used method for this purpose. Tang *et al.* [30] enhanced the target discrimination by target domain clustering and structural source domain regularization. Huang *et al.* [21] proposed a method for determining the likelihood that an input sample belongs to a cluster via a dynamic clustering confidence value. In [24], the authors designed an uncertainty-guided noise resilient network to mitigate the negative effects of noise pseudo labels when clustering.

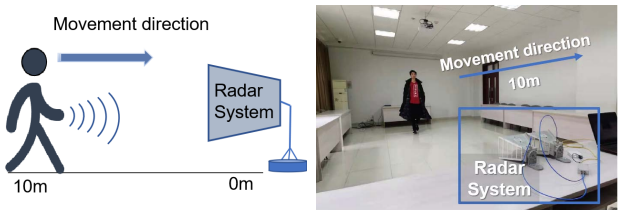
3) Subspace-Based UDA: To construct a domain-invariant subspace, using geometric constraint is also explored in these years. Gong *et al.* [40] employed numerous low-dimensional subspaces for establishing a domain-invariant flow path based on the geodesic flow kernel. The authors of [41] explored a feasible solution for constructing a discriminative domain-agnostic subspace by using the two-stage learning strategy.

The purpose of this work is to address the problem of person identification based on disguised gaits by offering a novel UDA-based method. Unlike the approaches described above, our method is more suitable to the concerned problem because we take the differences in the discrepancy between the source and target domains as a significant prior in the design of G-SAC.

III. Micro-Doppler Dataset

In this part, we first present a measured radar gait dataset of m-Ds to evaluate the proposed approach. To accomplish this, a SISO radar system with an ultra-wideband (UWB) radar module and two directional antennas is built. The frequency band of the radar system is 3.1–4.8 GHz, with a pulse repetition frequency of 368 Hz.

Eight gait patterns are considered in the measurements, which include: walking normally (**N**ormal), walking fast (**F**ast), walking slowly (**S**low), walking with a backpack (**B**ackpack), walking with a handbag (**W**ith bag), walking with



(a) (b)

Fig. 3. (a) The measurement setup. (b) Experimental environment in the 'Coat' disguised pattern.

TABLE I
THE NUMBER OF GENERATED SAMPLES OF RADAR MEASUREMENTS FOR EACH GAIT PATTERN

Domain	N	B	C	P	S	St	W
Samples	12239	12759	12520	13089	17557	12899	12626

a stick (**St**ick), walking with a cell phone (**P**hone), and walking in a coat (**C**oat). Nine subjects including four males and five females participated in the radar measurements. Subjects perform eight patterns starting from a distance of 10 m towards the radar (see Fig. 3). All of the experiments are operated in an indoor environment. Each participant performs a single gait pattern ten times, and each measurement is recorded as an echo segment. Thus, the total number of echo segments was (9persons) $\times$ (8patterns) $\times$ (10repeats) $=$ 720.

After the radar measurements, we obtain the m-Ds by following two steps. Firstly, in order to obtain the human backscattering echos with a high signal-to-noise ratio (SNR), we follow a standard pipeline to improve the quality of raw radar data. Average background subtraction is employed to remove the direct current (DC) component, and then we filter the echos from static objects with a moving target indicator (MTI), so that the component of human movements in raw radar signal can be completely retained. The threshold detection is the next operation, which is used for suppressing the noise in the raw radar signals. Finally, we may acquire radar echoes free of interference from the environment or electrical circuits.

The second step is to produce radar spectrograms using the time-frequency transform. We use the short-time Fourier transform (STFT) to convert radar signals to spectrograms in our study. The STFT is calculated based on a 0.2 s Hanning window and 2048 sampling points, which can be formulated as follows,

$$S_{\mathrm{STFT}}(\tau,\omega)=\left|\int_{-\infty}^{\infty}x(t-\tau)h(t)\mathrm{d}t\right|^2, \tag{1}$$

where $x(t)$ denotes signal segments in the time domain, and $h(t)$ is the window function. Afterward, we use the sliding window augmentation [33] on each echo segment, and set 90% overlap to divide the signal into segments and perform windowing. We finally obtain more than 90,000 spectrograms on all the measurements. The numbers of each gait pattern's images are listed in Table I.

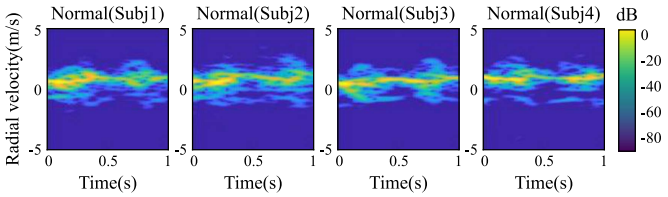


Fig. 4. The Micro-Doppler spectrograms of four subjects in 'Normal' pattern.

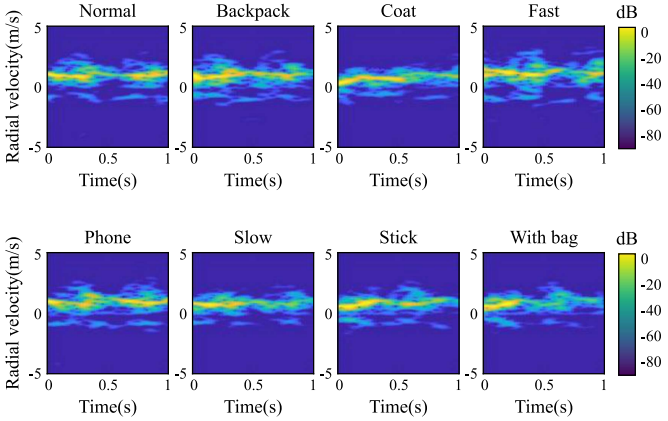


Fig. 5. The Micro-Doppler spectrograms of the same subject in eight gait patterns.

From Fig. 4, we can observe intuitively that the m-Ds of different targets with the same pattern have similar envelopes. Meanwhile, there are clear distinctions between the normal gaits and several disguised gaits (**C**, **S**, **St**, and **W**). The distinctions are probably caused by the change of Doppler shifts as shown in Fig. 5 or period of motion, and this dissimilarity might occur locally (different echo intensities) or globally (different Doppler bandwidth).

IV. PROPOSED METHOD

The G-SAC model is introduced in this section. To begin with, we formulate the disguised-gait-based person identification problem in Sec. IV-A. Then, the unsupervised deep clustering constraints are presented in Sec. IV-B. We detail the NCA-guided intrinsic subspace learning in Sec. IV-C. The following section presents the joint supervised and unsupervised optimization, and the proposed self-adaptive class-aware alignment constraint. The general framework of the G-SAC model is depicted in Fig. 6.

A. Problem Formulation

In UDA, given the spectrograms for the normal scenario as the source domain $\mathcal{S}$ with n_s labeled samples $\mathcal{X}^s = \left\{\left(x_i^s, y_i^s\right)\right\}_{i=1}^{n_s}$ and the spectrograms for one of the disguised scenarios as the target domain $\mathcal{T}$ with n_t unlabeled samples $\mathcal{X}^{\mathrm{t}} = \left\{\left(x_j^t\right)\right\}_{j=1}^{n_t}$, where $\mathcal{X}^s, \mathcal{X}^t \in \mathbb{R}^D$. Source domain $\mathcal{S}$ and target domain $\mathcal{T}$ share a label space $\mathcal{Y}$, where $y^{\mathcal{S}} \in \{1, 2, \ldots, K\}$ denote the source domain label of K classes. The target domain label $y^{\mathcal{T}} \in \{1, 2, \ldots, K\}$ cannot be

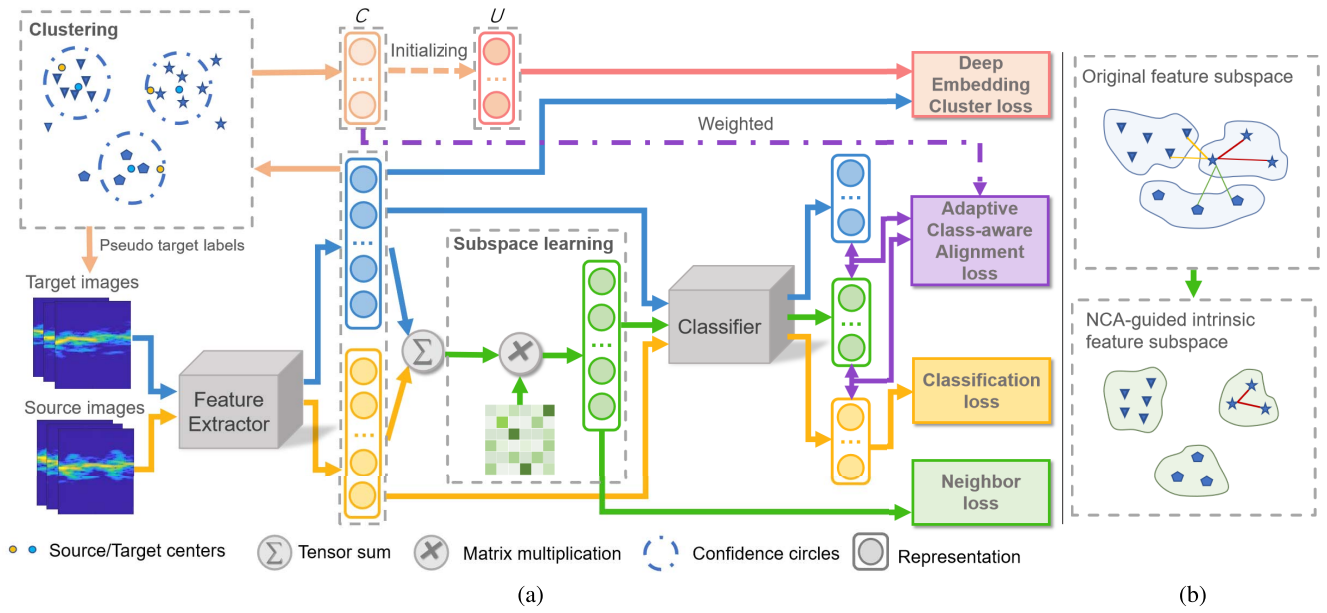


Fig. 6. (a) The proposed G-SAC framework is an end-to-end network shared by the source and target domains. In the subspace learning phase, the feature extractor captures the original feature representation, followed by using NCA to guide the subspace learning and finally using the classifier to predict. In the clustering phase, we use standard K-means clustering based on the current feature representation of source and target samples. (b) Visualization of the NCA-guided subspace learning process. The network projects the original features onto the intrinsic subspace by learning a set of mappings with parameters. Detailed description can be found in Section IV.

accessed. UDA aims to make the prediction of each $\hat{y}_j^t$ using $\mathcal{X}^s$, $\mathcal{X}^t$, and $\mathcal{Y}^s$.

Our approach is to bridge the gap in the distribution of the source and target domains with a neural network. The G-SAC consists of three modules: a feature extractor $f_e(\cdot;\theta)$, a classifier $f_c(\cdot;\phi)$ and a guiding module $f_g(\cdot;\omega)$, where θ, ϕ and ω are the parameters of the corresponding modules respectively. Given the feature extraction network $f_e(\cdot;\theta)$, the inputs $x \in \mathcal{X}$ are mapped into a raw feature space $\mathcal{Z} \in \mathbb{R}^d$ ($d \ll D$) by a feature embedding process: $\mathcal{X} \xrightarrow{\theta} \mathcal{Z}$. Let $f_g : \mathcal{Z} \xrightarrow{\omega} \mathcal{O}$ be a transformation from the input $z \in \mathcal{Z}$ to the intrinsic subspace $\mathcal{O} \in \mathbb{R}^d$. The classifier $f_c(\cdot;\phi)$ assigns a probability vector $p \in \mathbb{R}^K$ to each feature representation $z \in \mathcal{Z}$ or $o \in \mathcal{O}$ by seeking an optimal h in the functional space $\mathcal{F}_h$, where $h \in \mathcal{F}_h : \{\mathcal{Z} \xrightarrow{\phi} \mathbb{R}^K \;\&\&\; \mathcal{O} \xrightarrow{\phi} \mathbb{R}^K\}$, and the optimization is constrained by $p_{i,k}^s = y_i^s$.

B. Deep Target Domain Discriminative Clustering

In our design, the clustering is implemented in the original feature space, so we choose a flexible embedding clustering [42] framework for deep clustering, as this method can take the feature representation as inputs. Each x_k^t is projected into the original feature space based on a backbone network $f_e(\cdot;\theta)$, and the feature representation $\{z_k^t\}_{k=1}^K$ is obtained. Next, we employ the unsupervised deep embedding cluster loss [42] to force $\{z_k^t\}_{k=1}^K$ to surround a group of learnable prototypes $U = \{\mu_k^t, k = 1, 2, \ldots, K\}$ in terms of k-th class.

The objective function of the deep target domain discriminative clustering is defined as

$$\min_{\theta, Q^t, \{\mu_k^t\}_{k=1}^K} \mathcal{L}_{cluster}^{\mathrm{t}} = \mathrm{KL}(Q^t \parallel \tilde{P}^t) - H(\rho), \tag{2}$$

where $\tilde{P}^t = \{\{\tilde{p}_{j,k}^t\}_{j=1}^{n_k^t}\}_{k=1}^K$ denotes the probability distributions of z^t and μ^t, $Q^t = \{\{q_{j,k}^t\}_{j=1}^{n_k^t}\}_{k=1}^K$ is the auxiliary target distribution evolving from $\tilde{P}^t$, $H(\rho)$ refers to the entropy loss of ρ which is the expectation of Q^t, and the KL divergence loss can be written as

$$\mathrm{KL}(Q^t \parallel \tilde{P}^t) = \frac{1}{n_t} \sum_{j=1}^{n_t} \sum_{k=1}^{K} q_{j,k}^t \log\left(\frac{q_{j,k}^t}{\tilde{p}_{j,k}^t}\right). \tag{3}$$

We use the Student's t-distribution [43] combined with a multinomial logistic regression (softmax) function to calculate the probability $\tilde{p}_{j,k}^t$ of assigning z_j^t to μ_k^t in the target domain, which is

$$\tilde{p}_{j,k}^t = \frac{\exp\left(\left(1 + \left\|z_j^t - \mu_k^t\right\|^2/\delta\right)^{-\frac{\delta+1}{2}}\right)}{\sum_{k'=1}^{K} \exp\left(\left(1 + \left\|z_j^t - \mu_{k'}^t\right\|^2/\delta\right)^{-\frac{\delta+1}{2}}\right)}. \tag{4}$$

Following [42], we let $\delta = 1$ for all experiments.

Similar to [30], [42], the $\tilde{P}^t$ is iteratively refined by learning from high confidence assignments with the help of Q^t. To update the auxiliary target distribution Q^t, an alternative update strategy is employed. At the first stage, $f_e(\cdot;\theta)$ and $\{\mu_k^t\}_{k=1}^K$ are fixed. We can get the closed-form solution of $q_{j,k}^t$

with the subject to $\sum_{k=1}^{K} q_{j,k}^{t} = 1$ [44],

$$q_{j,k}^{t} = \frac{\tilde{p}_{j,k}^{t} / \sqrt{\sum_{j'=1}^{n_t} \tilde{p}_{j',k}^{t}}}{\sum_{k'=1}^{K} \tilde{p}_{j,k'}^{t} / \sqrt{\sum_{j'=1}^{n_t} \tilde{p}_{j',k'}^{t}}}. \tag{5}$$

Next, we fix Q^t and update $f_e(\cdot;\theta)$ and $\{\mu_k^t\}_{k=1}^{K}$ following Eq. (6), i.e.,

$$\min_{\theta,\{\mu_k^t\}_{k=1}^{K}} \mathcal{L}_{cluster}^{\mathrm{t}} = -\frac{1}{n_t} \sum_{j=1}^{n_t} \sum_{k=1}^{K} q_{j,k}^{t} \log \tilde{p}_{j,k}^{t}, \tag{6}$$

which can be treated as a standard cross-entropy loss function with the target distribution Q^t as the label.

As stated in [45], degenerate solutions probably occur when excessive instances are assigned to the same cluster. We introduce a regularization term $H(\rho)$ to balance the number of assignments in clusters. This term $H(\rho) = -\sum_{k=1}^{K} \rho_k \log \rho_k$ denotes the entropy of the probability $\rho_k = \frac{1}{n_t} \sum_{j=1}^{n_t} q_{j,k}^{t}$. This objective function is encouraged to maximize the entropy of Q^t, so that the total loss can equally contribute to each cluster. In addition, this can allow the cluster boundaries of the target domain to pass through low-density regions [35].

In addition to serving as the unsupervised constraints of the G-SAC, the clustering is also assigned the task of predicting a pseudo-label for each x_j^t which is required for the NCA-guided subspace modeling (in Sec. IV-C), and the detailed procedure of generating pseudo labels is presented in Sec. IV-D.

C. NCA-Guided Intrinsic Subspace Learning

The study in [30] notes that the source generalization error will be accumulated when the domain alignment does not take the class-level information into consideration at an early stage. Based on this study, we take a closer perspective from domain-level to class-level. Specifically, we construct a learnable *class-aware intrinsic subspace* and introduce an alignment constraint to alleviate the risk of discrepancies in the degree between normal and various disguised gaits.

1) Neighbourhood Component Analysis Overview: NCA [31] aims to find the class-aware feature representations using a linear/non-linear transformation. Its appeal comes from the fact that the features can preserve their manifold structure after the transformation. Compared with the classical PCA [46] and linear discriminant analysis (LDA) [47] algorithms, NCA neither requires complicated matrix inversion operations nor does it need to make strong parametric assumptions on the class distribution structure and decision boundaries [35].

NCA is closely related to the K-nearest neighbor (KNN) algorithm and handles the tasks of feature selection and dimensionality reduction with this algorithm. Because the NCA takes the distance differences among each pair of samples into account based on a KNN-determined metric space, it is more appropriate for the problem addressed in this paper.

2) NCA-Guided Subspace Modeling: Given a source domain feature embedding z^s and a target domain feature embedding z^t. We pre-mix z^s and z^t for each class k before NCA-based projection. The mixing operation follows Eq. (7),

$$\tilde{z}_{k,ij}^{st} = \beta z_{k,i}^{s} + (1-\beta) z_{k,j}^{t}, \tag{7}$$

where $\beta \in [0,1]$ is a trade-off coefficient to adjust the mixing ratio of the source and target domain features. A parametric metric based on Mahalanobis distance is employed to project the mixed features into an intrinsic feature subspace, and this step is strictly restricted by the NCA assumption. The Mahalanobis distance is commonly used in cluster analysis and classification approaches due to its advantages of being unitless, scale-invariant, and taking into account the dataset's correlations [48]. In terms of the NCA principal, the parametric metric should be considered as a positive semi-definite matrix, and given the constraint of $\mathcal{Q} = \omega^T \omega$. This metric can be formulated as:

$$\begin{aligned} \mathbb{M}(m,n) &= (m-n)^T \mathcal{Q} (m-n) \\ &= (\omega m - \omega n)^T (\omega m - \omega n). \end{aligned} \tag{8}$$

According to Sec. IV-A, we can obtain $\mathcal{O}^{st}$ from the mixed feature subspace $\tilde{\mathcal{Z}}^{st}$, and the distances between each random pairs in $\mathcal{O}^{st}$ can be calculated by Eq. (8) following the random neighborhood soft assignment strategy [31]. Then, we can get the probability q_{mn} of assigning each mixed feature $\tilde{z}_m^{st}$ to $\tilde{z}_n^{st}$ $(m \neq n)$ using the softmax operator in subspace $\tilde{\mathcal{Z}}^{st}$, which is

$$q_{mn} = \frac{\exp\left(-\mathbb{M}(\tilde{z}_m^{st}, \tilde{z}_n^{st})\right)}{\sum_{\substack{n=1 \\ m \neq n}}^{n_{st}} \exp\left(-\mathbb{M}(\tilde{z}_m^{st}, \tilde{z}_n^{st})\right)}, \tag{9}$$

where n_{st} denotes the number of intrinsic subspace samples. Finally, we can get the probability of correct assignments of $\tilde{z}_m^{st}$ following

$$p_m = \sum \mathbb{I}(k_m = k_n)\, q_{mn}, \tag{10}$$

where the $\mathbb{I}(\cdot)$ is the indicator function. It should be noted that a few numbers of wrong pseudo-labels would not affect the optimization, because 1) the p_m is determined by all the q_{mn} where the $k_m = k_n$ condition is satisfied, and 2) each $\tilde{z}_m^{st}$ is derived from the subspace $\mathcal{Z}^s$ where the samples are accurately labeled.

In order to avoid the risk of overfitting, an additional regularization term is considered to constrain the parameter space of the learnable parameters ω based on the structural risk minimization (SRM) theory [49]. Based on the consideration above, the objective function for establishing the NCA-guided intrinsic subspace is defined as follows:

$$\min_{\omega} \mathcal{L}_{neighbor}^{st} = -\frac{1}{n_{st}} \sum_{m}^{n_{st}} \log(p_m) + \tau \|\omega\|_F^2, \tag{11}$$

where τ acts as a non-negative regularization parameter, and $\|\cdot\|_F^2$ indicates the Frobenius matrix norm. This term ensures complexity control on ω elements, so that the high-dimensional intrinsic features in subspace $\mathcal{O}^{st}$ would become more robust in extracting discriminative features [50].

3) Class-Aware Alignment Constraint: The class-aware alignment loss is designed to align the subspace $\mathcal{Z}$ with the subspace $\mathcal{O}^{st}$ in both source and target domain conditions, so that the features in $\mathcal{Z}$ would be class-aware and domain-invariant. The class-aware alignment loss is formulated as follows,

$$\begin{aligned}\min_{\theta,\phi} \mathcal{L}^{st}_{\text{align}} &= \underbrace{\frac{1}{K}\sum_{k=1}^{K}\left(\mathcal{L}_{wd}\left(\mathcal{Z}^s_k, \mathcal{O}^{st}_k\right) + \mathcal{L}_{wd}\left(\mathcal{Z}^t_k, \mathcal{O}^{st}_k\right)\right)}_{intra} \\ &\quad \underbrace{- \frac{1}{K}\frac{1}{K-1}\sum_{k=1}^{K}\sum_{\substack{k'=1 \\ k'\neq k}}^{K}\mathcal{L}_{wd}\left(\mathcal{O}^{st}_k, \mathcal{O}^{st}_{k'}\right)}_{inter}, \end{aligned} \tag{12}$$

where $\mathcal{L}_{wd}$ is the Wasserstein distance (also called The Earth Mover Distance). Following [37], we simply calculate the Wasserstein distance under the Kantorovich-Rubinstein dual principle, i.e.,

$$\mathbb{WD}\left(\mathcal{X}^s, \mathcal{X}^t\right) = \sup_{\left\|\hat{f}\right\|_L \leq 1} {}_{x^s\sim\mathcal{X}^s}\left[\hat{f}\left(x^s\right)\right] - {}_{x^t\sim\mathcal{X}^t}\left[\hat{f}\left(x^t\right)\right], \tag{13}$$

where $\left\|\hat{f}\right\|_L$ is the Lipschitz constant for $\hat{f}(\cdot)$. In this case, the $\mathcal{L}_{wd}$ in Eq. (12) can be formulated as

$$\mathcal{L}_{wd}\left(\mathcal{Z}_k, \mathcal{O}^{st}_k\right) = \frac{1}{n_k}\sum_{z_k\in\mathcal{Z}_k} f_c\left(z_k\right) - \frac{1}{n^{st}_k}\sum_{o^{st}_k\in\mathcal{O}^{st}_k} f_c\left(o^{st}_k\right) \tag{14}$$

and the z_k can be z^s_k or z^t_k in the training phase.

It should be pointed out that the intra-class terms of Eq. (12) aim at narrowing the domain gap by minimizing the distances between the intrinsic features and original features in both the source and target domains, while also reducing the intra-class discrepancies between those features. The inter-class term mainly focuses on increasing the dissimilarity between different classes in the subspace $\mathcal{O}^{st}$ by maximizing the distances between each pair of intrinsic features from different classes. Note that the Wasserstein distance is calculated based on the expectations of distributions, so it is considered inherently noise-robust against the few incorrect pseudo-labels [37].

Combining Eq. (11) and Eq. (12), we can obtain the total objective function for the NCA-guided intrinsic subspace learning:

$$\min_{\theta,\phi,\omega} \mathcal{L}^{st}_{\text{guide}} = \mathcal{L}^{st}_{\text{align}} + \mathcal{L}^{st}_{\text{neighbor}}. \tag{15}$$

Besides, we minimize the classification error of the source domain labeled data by using the cross-entropy loss

$$\min_{\theta,\phi} \mathcal{L}^{s}_{\text{cls}} = -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{k=1}^{K}\mathbb{I}\left[k = \text{y}^s_i\right]\log p^s_{i,k}, \tag{16}$$

where $p^s_{i,k}$ is the k-th element of the predicted probability vector p^s_i of source instance x^s_i. Therefore, the overall objective of G-SAC can be formed as

$$\min_{\theta,\phi,\omega,Q^t,\{\mu^t_k\}_{k=1}^K} \mathcal{L}_{\text{total}} = \mathcal{L}^s_{\text{cls}} + \alpha\mathcal{L}^t_{\text{cluster}} + \gamma\,\mathcal{L}^{st}_{guide}. \tag{17}$$

D. Joint Supervised and Unsupervised Optimization

In each epoch of training, the optimization of the G-SAC model and the clustering in the original feature subspace $\mathcal{Z}$ are performed alternately. Clustering is crucial in our approach since it enables the combination of supervised and unsupervised optimization. For the supervised learning, the clustering method provides the pseudo labels of each x^t_j, and the $\mathcal{L}^{st}_{guide}$ is heavily associated with these pseudo labels. Meanwhile, the $\mathcal{L}^t_{\text{cluster}}$ is affected by the learnable prototypes μ^t, which are acquired from the cluster centers c^t provided by the clustering method (the dash line in Fig. 6). It is noted that c^t differs from μ^t as the former is initialized by the clustering at the beginning of each epoch, while the latter is continuously updated under the constraint of Eq. (2).

1) Cluster Center Updating: In our design, as the cluster centers have a substantial impact on the pseudo labels, the assignment and optimization of these cluster centers are detailed in this section.

At the start of each epoch, we employ the K-means [51] algorithm to assign pseudo-labels to each x^t_j by taking two factors, one is the number of categories K in the source domain $\mathcal{S}$, the other is the cluster centers of the samples $\mathcal{X}^s$ to initialize the c^t_k: $c^t_k = c^s_k$, where c^s_k is computed by $c^{\text{s}}_k = \frac{1}{n^{\text{s}}_k}\sum_{z^{\text{s}}_k\in\mathcal{Z}^{\text{s}}_k}\frac{z^{\text{s}}_k}{\|z^{\text{s}}_k\|}$.

Then, the cluster centers c^t are further optimized based on the error minimization algorithm [51] following two steps. The first step is to calculate the vector $v_{z^t_j\to c^t} = \left[d_{\cos}\left(z^t_j, c^t_1\right), d_{\cos}\left(z^t_j, c^t_2\right), \ldots, d_{\cos}\left(z^t_j, c^t_K\right)\right]$ based on the cosine similarity metric $d_{\cos}(\cdot)$. In this case, the index number of the maximum value in $v_{z^t_j\to c^t}$ is considered as the pseudo label of the input z^t_j. The second step is to re-assign pseudo label of z^t_j, and the c^t_k is updated until the error between each re-assignment is less than an extremely low number ϵ.

After clustering, we set a threshold $D_0 \in (0, 1)$ to filter those features far away from each c^t_k, i.e., $\left\{\left(x^t_j, \hat{y}^t_j\right) | d_{\cos}\left(z^t_{j,k}, c^t_k\right) > D_0\right\}$. This step seeks to improve the pseudo label prediction results by rejecting those ambiguous cases.

2) Self-Adaptive Alignment Weights: As mentioned in the introduction, there are discrepancies between the two data domains. Furthermore, another issue arises from this observation: the disparities between inter-class and intra-class samples with different domains are likewise diverse. In order to deal with the problem of this imbalance within each class and between different classes, we design two self-adaptive weights for Eq. (12) base on the cluster centers c^s, c^t and the cosine similarity distance [30]. The k-th intra-class weight is calculated by $w^{kk}_{\text{intra}} = 1 - \frac{c^{t\top}_k c^s_k}{\|c^t_k\|\|c^s_k\|} \in [0, 1]$, and the inter-class weight between class k and k' is obtained following

TABLE II
CLASSIFICATION ACCURACY (%) FOR DIFFERENT ADAPTATION TASKS ON THE MEASURED RADAR MICRO-DOPPLER DATASET

Method	N→B	N→C	N→F	N→P	N→S	N→St	N→W	Average
Source-only [52]	66.01	38.26	51.79	59.72	47.94	49.79	48.84	51.76
DAAN [27]	69.03	40.66	71.40	62.27	44.03	51.23	51.10	55.67
DANN [38]	91.18	57.15	81.18	74.72	61.90	68.93	66.37	70.20
BNM [22]	74.65	48.79	78.57	67.96	43.68	59.66	57.32	61.52
SAFN [39]	78.08	51.60	79.62	72.51	54.30	63.37	60.57	65.72
CLAN [26]	84.69	56.26	82.55	75.58	57.39	64.27	61.52	68.89
DCAN [35]	79.15	51.51	83.20	77.29	**68.10**	62.49	63.08	69.26
GVB [28]	83.60	49.57	85.63	80.14	59.33	66.78	62.67	69.67
DSAN [23]	84.94	56.27	86.62	79.01	53.88	68.72	65.96	70.77
SRDC [30]	83.48	59.79	81.41	73.84	50.30	76.89	72.41	71.16
ATDOC [53]	83.23	51.49	81.82	78.52	64.14	67.53	59.88	69.52
FixBi [54]	84.47	63.07	85.16	81.48	61.94	75.43	65.73	73.90
G-SAC	**89.96**	**63.34**	**94.95**	**88.07**	65.87	**82.28**	**77.86**	**80.33**

$w_{\text{inter}}^{kk'} = \frac{(c_k^t + c_k^s)^\top (c_{k'}^t + c_{k'}^s)}{\|c_k^t + c_k^s\| \left\|c_{k'}^t + c_{k'}^s\right\|} \in [0, 1]$. The weighted alignment loss is given by

$$\begin{aligned}\min_{\theta,\phi,\omega} \mathcal{L}_{\text{align}}^{st} &= \underbrace{\frac{1}{K}\sum_{k=1}^{K} w_{\text{intra}}^{kk}\left(\mathcal{L}_{wd}\left(\mathcal{Z}_k^s, \mathcal{O}_k^{st}\right) + \mathcal{L}_{wd}\left(\mathcal{Z}_k^t, \mathcal{O}_k^{st}\right)\right)}_{intra} \\ &\quad - \underbrace{\frac{1}{K}\frac{1}{K-1}\sum_{k=1}^{K}\sum_{\substack{k'=1 \\ k'\neq k}}^{K} w_{\text{inter}}^{kk'}\mathcal{L}_{wd}\left(\mathcal{O}_k^{st}, \mathcal{O}_{k'}^{st}\right)}_{inter}. \end{aligned} \tag{18}$$

By incorporating weights into the objective function, it is possible to perform the class-aware alignment procedure using the clustering results as a prior. In this case, we could pay different attention to each class, and the class-level disequilibrium within the features in subspace $\mathcal{O}^{st}$ and $\mathcal{Z}$ will be eliminated.

V. EXPERIMENTAL RESULTS

A. Setups

Typically, the UDA problem is seen as a single-target domain adaptation (STDA) problem, with identical classes in the source and target domains. By expanding the number of target domains, a more complex UDA problem, the multi-target domain UDA (MTDA) problem, was recently introduced [55].

1) Dataset Construction: We assign the **N** domain as the source domain based on the measured radar data mentioned in Sec. III, and the other domains as the target domains independently. We build eight multi-target domains in the MTDA setup by randomly selecting two domains from among those that are not the **N** domain base on the measured radar data.

2) Implementation Details: The input spectrograms are augmented by random cropping and horizontal flipping, and each of them is resized to 224 × 224 pixels before being fed into the network. The ResNet-50 [52] with the ImageNet pre-trained model [56] is selected as our backbone network (the feature extractor $f_e(\cdot;\theta)$ in Sec. IV-A), and the last fully-connected (FC) layer is removed. The classifier ($f_c(\cdot;\phi)$ in Sec. IV-A) is a 3-layer perceptron. The mini-batch stochastic gradient descent (SGD) solver is used for the optimization, and its momentum is 0.9 and weight decay is 0.001. We set the initialized learning rate at 0.001 for $f_e(\cdot;\theta)$, 0.01 for $f_c(\cdot;\phi)$, and 0.0001 for $f_g(\cdot;\omega)$. We follow the same learning rate strategy as [30]: the learning rate decays with the factor of $\left(1+\eta\frac{e}{E}\right)^{-\rho}$, where e and E denote the process of training epochs and total epochs respectively, $\eta = 10$, and $\rho = 0.75$. We follow [27] to increase γ from 0 to 1 by $\gamma_e = 2 \times (1+\exp(-\vartheta \mathrm{e}))^{-1} - 1$, where $\vartheta = 10$. We set the mixing factor $\beta = 0.5$ (see Sec. IV-C) and the filtering threshold $D_0 = 0.05$ (see Sec. IV-D). We follow [50] to set $\tau = 27$ (see Sec. IV-C). The maximum number of epochs is set to 100, and the batch size is set to 16. Our implementation is based on the PyTorch open-source framework [57].

B. Results on Radar-Based Person Identification Task

The state-of-the-art UDA algorithms, including the domain-alignment-based methods BNM [22], SAFN [39], and DCAN [35], adversarial-learning-based methods DANN [38], DAAN [27], and GVB [28], class-aware-based methods SRDC [30], CLAN [26], DSAN [23], and pseudo-label-based methods ATDOC [53] and FixBi [54] are taken into consideration for the comparison.

1) STDA Setup: The results of all approaches on the proposed dataset in the STDA setup are shown in Table II. Based on the above source-only results, we group **C**, **S**, **St** and **W** with less than 50% accuracy as *strongly-disguised* gaits, whereas **B**, **F**, and **P** as *weakly-disguised* gaits. As demonstrated in the table, our method outperforms the other methods by at least 6.91%, proving our approach's significant competitiveness for disguised-gait-based person identification. On the other hand, the significant disparities between the various accuracies (e.g., the accuracy of N→C is 63.34%, whereas the accuracy of N→F achieves 94.95%) demonstrate the variations across the UDA tasks. We find that DCAN achieves an excellent result in N→S, but it is not robust

TABLE III
CLASSIFICATION ACCURACY (%) FOR DIFFERENT MTDA TASKS OF THE MEASURED RADAR MICRO-DOPPLER DATASET

Method	N→BF	N→BP	N→FSt	N→WSt	N→PW	N→FP	N→BW	N→CP	Average
Source-only [52]	60.28	63.52	50.68	49.56	53.69	55.32	58.63	48.51	55.02
GVB [28]	65.05	75.58	53.11	56.62	67.03	62.65	68.90	59.06	63.50
SRDC [30]	68.23	72.07	58.81	56.06	63.06	69.32	64.56	59.68	63.97
SAFN [39]	73.69	68.34	63.86	58.14	62.44	66.06	65.64	56.45	64.33
DCAN [35]	81.66	77.10	68.45	56.11	64.05	74.72	65.68	61.65	68.68
CLAN [26]	80.00	79.42	71.82	64.87	72.29	77.30	71.83	65.19	72.84
ATDOC [53]	80.58	74.73	67.39	60.56	68.49	75.57	72.01	61.70	70.13
FixBi [54]	84.92	86.04	74.25	68.84	74.49	82.42	78.01	66.77	76.97
G-SAC(ours)	**92.19**	**88.79**	**82.77**	**72.74**	**77.95**	**87.24**	**81.21**	**71.16**	**81.76**

TABLE IV
CLASSIFICATION ACCURACY (%) OF ABLATION STUDIES ON THE MEASURED RADAR MICRO-DOPPLER DATASET. FOR DEFINITIONS OF THE DIFFERENT METHODS, SEE THE TEXT IN SECTION IV

Method	N→B	N→C	N→F	N→P	N→S	N→St	N→W	Average
Source-only	66.01	38.26	51.79	59.72	47.94	49.79	48.84	51.76
G-SAC(w/o NCA-guided intrinsic subspace learning)	83.47	52.71	86.77	84.61	54.53	69.41	66.01	71.07
G-SAC(w/o deep embedding distinguishing clustering)	84.67	55.78	83.84	82.12	50.27	70.75	63.37	70.11
G-SAC(w/o self-adaptive alignment weights)	89.79	55.94	93.27	86.18	57.93	74.71	70.25	75.44
G-SAC	**89.96**	**63.34**	**94.95**	**88.07**	**65.87**	**82.28**	**77.86**	**80.33**

enough in *weakly-disguised* gaits, and GVB can identify the target accurately in *weakly-disguised* gaits, but it cannot fight the disturbance in *strongly-disguised* gaits. By contrast, our method can effectively compensate for large domain discrepancy while avoiding over-transfer for small domain discrepancy and achieves robust performance in any disguised gait condition.

2) MTDA Setup: Table III shows the comparison results on the radar micro-Doppler dataset under the MTDA setup. The results indicate that our method achieves the SOTA performance for the disguised-gait-based person identification task.

C. Ablation Studies

1) Effectiveness of NCA-Guided Intrinsic Subspace Learning: To verify the efficacy of the intrinsic subspace learning, we remove the $\mathcal{L}_{guide}^{st}$ in G-SAC and perform the experiments on this variant based on the dataset in Sec. III. The ablation results in Table IV demonstrate the important role of intrinsic subspace learning in the problem at hand. The average accuracy reduces by nearly 10% when compared to the G-SAC. In this task, considering the large number of biomechanical gait variables involving various joints and planes of motion, the results indicate that the NCA focuses on the information concerning individual discrimination, while discarding the irrelevant pieces of information.

2) Effectiveness of Deep Target Domain Discriminative Clustering: The $\mathcal{L}_{\text{cluster}}^{t}$ is removed to verify the effectiveness of clustering. In Table IV, we note that clustering has a significant impact on improving the accuracy, demonstrating the critical relevance of giving pseudo labels during the feature extractor's learning process.

3) Effectiveness of Self-Adaptive Alignment Weights: We conducted the ablation experiments by replacing the Eq. (18) with Eq. (12) in the G-SAC model, and the corresponding results on the seven adaptation tasks are listed in Table IV. As can be seen, the proposed self-adaptive weights are beneficial in all cases. Specifically, accuracy increases significantly for scenarios in which the target domain is obtained using *strongly-disguised* gaits, however, there is little difference between the counterparts of *weakly-disguised* gaits scenarios. This observation also suggests that the class-level imbalance is more severe in scenarios with strongly disguised gaits than in those with weakly disguised gaits.

4) Effectiveness of the Class Alignment With Wasserstein Distance: We additionally discuss the efficacy of the Wasserstein distance in the class-level alignment. We chose some commonly used metrics in transfer learning, including MMD [35] and KL divergence [36], to replace the Wasserstein distance employed in this paper, and the results are depicted in Fig. 7a and Fig. 7b. It can be observed in these figures that the accuracy curve of the proposed method is smoother than those using MMD and KL divergence. Moreover, G-SAC performs more accurately than the other variations in those challenging adaptation tasks involving strongly disguised gaits, e.g., the N→S.

D. Parameters Study

This section discusses the effects of pre-defined parameters on the performance of the G-SAC model, including the clustering loss weight α and the mixing factor β. As it can be seen in Fig. 7c, the accuracy fluctuates only in a small range when $\alpha \in \{0.5, 0.75, 1\}$. The results prove that our model is not susceptible to the change of α. We set the range

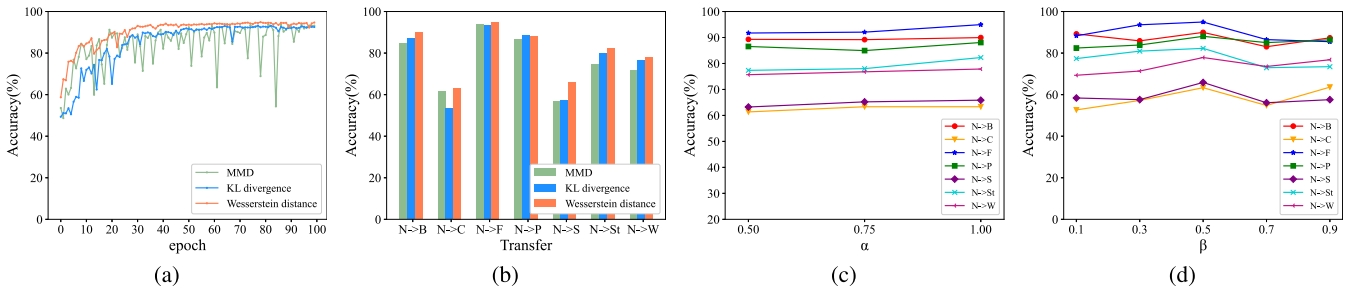


Fig. 7. (a) and (b) show comparison results of our method and two variants. (c) and (d) show accuracy (%) under different values of different parameters α and β.

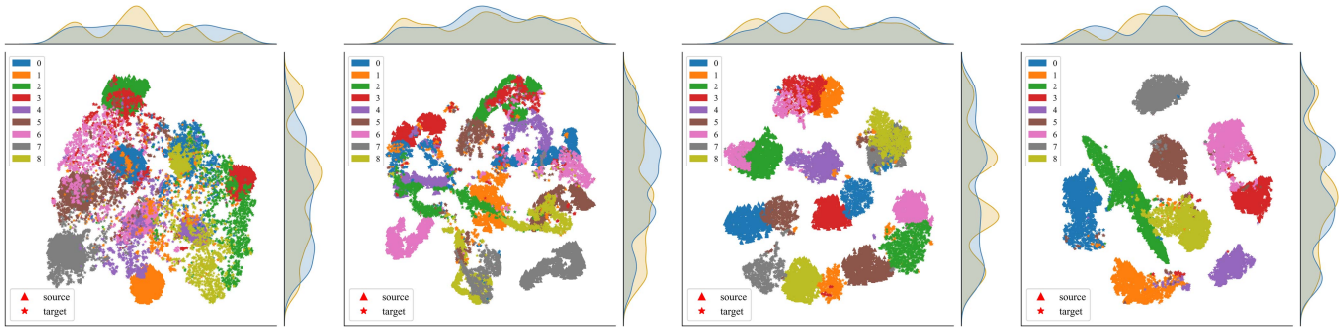


Fig. 8. The t-SNE visualization of embedded features on the transfer task N→F, where the numbers in the legend indicate nine human subjects. Note that the top and side plots represent the marginal domain distributions (where the source domain is in yellow and the target domain is in blue), while the middle plots represent the class conditional distributions.

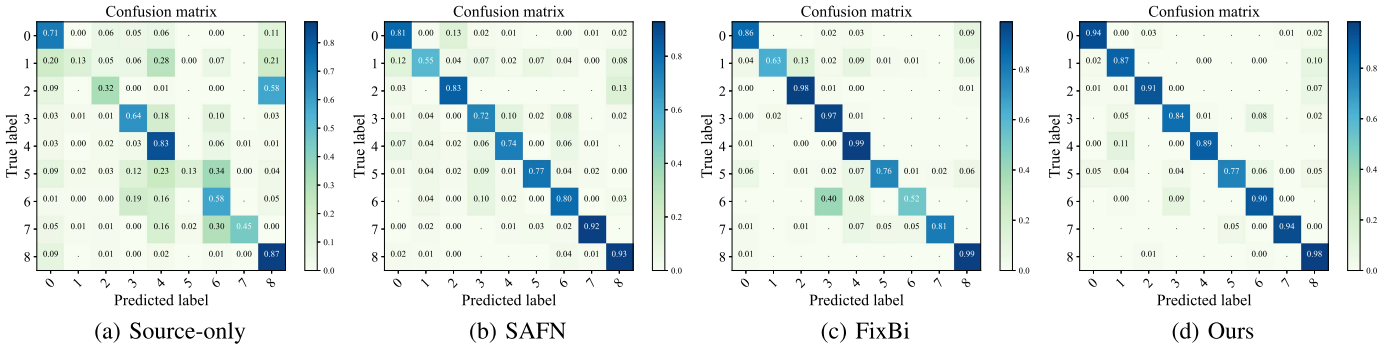


Fig. 9. The target domain confusion matrices on the transfer task N→F.

of $\beta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, as the corresponding results are shown in Fig. 7d. The best average accuracy is achieved when $\beta = 0.5$. This intriguing result reveals an interesting observation: a large β results in a weak connection between the intrinsic subspace and the target domain, whereas a small β results in a lack of supervision from the source domain. Thus, a better result can be achieved by averaging the contributions of both the source and target domains.

E. Further Discussion

1) Feature Visualization: We use the t-SNE [43] algorithm to visualize the distribution of features for the classification based on the dataset in Sec. III. Fig. 8 illustrates the feature distribution in the 2-dim space based on the adaptation task N→F. Our approach, as expected, generates class-discrepancy features in the target domain. Within the same class, there is still a distinct domain boundary between the source and target domains, demonstrating that the domain-level data structure can be well maintained.

2) Confusion Matrix: Fig. 9 shows the confusion matrix for the misclassification analysis based on the N→F adaptation task. From the confusion matrix of our method shown in Fig. 9d, we can see that the rate of misclassification between subjects is quite low. We can also see that the No.3 and No.6 subjects are prone to confusion on this adaptation task, as the misclassification rates are greater than 10% in all other compared algorithms. This reaches up to 58% when the classification model is only trained by the source data. Under comparison, our algorithm achieves a misclassification rate of

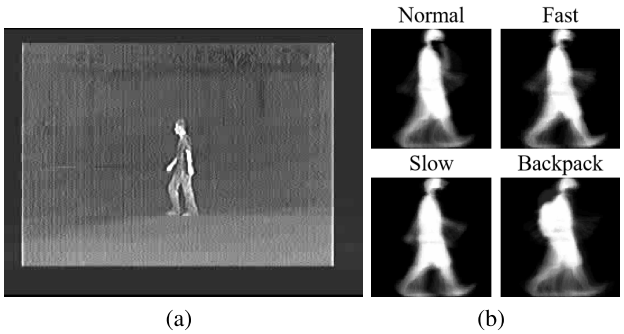


Fig. 10. (a) CASIA-C data collection. (b) GEI samples.

TABLE V
CLASSIFICATION ACCURACY (%) FOR DIFFERENT ADAPTATION TASKS ON CASIA-C DATASET

Method	N→F	N→B	N→S	Average
Source-only [52]	83.48	49.82	83.34	72.21
DAAN [27]	87.83	51.35	85.87	75.02
DANN [38]	90.60	85.31	87.77	87.89
SAFN [39]	84.44	65.72	84.12	78.09
CLAN [26]	91.10	84.85	89.45	88.47
BNM [22]	86.36	75.52	90.28	84.05
GVB [28]	90.76	77.97	85.00	84.58
DSAN [23]	90.43	78.71	88.95	86.03
SRDC [30]	89.56	79.39	88.17	85.71
ATDOC [53]	92.31	89.66	90.59	90.85
FixBi [54]	91.69	87.43	90.68	89.93
G-SAC	**93.15**	**90.64**	**91.06**	**91.62**

7% in the identical condition. Moreover, from Figs. 9a and 9b, we can see that there is significant confusion between all targets with each other, indicating the UDA method's difficulties in handling such disguised-gait-based person recognition tasks.

F. Results on Video-Based Person Identification Task

We compare the performance of G-SAC with other approaches using a publicly accessible gait dataset, CASIA-C [58], to explore its effectiveness on the video-based person identification task. This dataset contains 1530 video segments that were captured using thermal infrared cameras with a resolution of 320 × 240 pixels and a frame rate of 25 frames per second (FPS). Each video segment depicts a single individual walking in one of four distinct ways (see Fig. 10a), including walking normally (**N**ormal), walking fast (**F**ast), walking slowly (**S**low), and walking with a backpack (**B**ackpack). There are 130 male participants and 23 female participants in total. The authors of this dataset also provide a representation method of human gait in video sequences, named gait energy image (GEI), which is shown in Fig. 10b.

The per-task adaptation performance can be found in Table V. From this Figure, it is clearly that our model surpasses the others on each adaptation task. Compared to the performance of the source-only model, we achieve a 19.41% improvement on the average accuracy. Furthermore, the G-SAC shows significant improvements in all tasks, especially in the most difficult N→B task.

VI. CONCLUSION AND FUTURE WORK

We address the challenge of person identification using disguised gaits in this research. We formulate this challenge as an unsupervised domain adaptation problem and propose a new model for solving it named 'G-SAC'. This approach projects both source and target data, i.e., normal and disguised gaits, into an intrinsic feature subspace for feature alignment, and uses supervised and unsupervised constraints to ensure the consistency of the class-aware data distributions. We create a micro-Doppler gait dataset to evaluate the G-SAC model, and the findings demonstrate that our method outperforms other existing approaches in both single- and multi-target domain adaptation scenarios. Additionally, ablation experiments and specific insights are included to demonstrate the usefulness of our designs. We further validate the performance of G-SAC on a publicly accessible CASIA-C dataset, and the results show that the proposed model could be employed successfully for video-based gait recognition tasks. Given that an individual's m-Ds are susceptible to the presence of other moving objects, future studies will focus on person identification tasks in scenarios involving several individuals or other moving subjects.

REFERENCES

[1] J. Wei, Y. Wang, Y. Li, R. He, and Z. Sun, "Cross-spectral iris recognition by learning device-specific band," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Oct. 4, 2021, doi: 10.1109/TCSVT.2021.3117291.
[2] D. M. Uliyan, S. Sadeghi, and H. A. Jalab, "Anti-spoofing method for fingerprint recognition using patch based deep learning machine," *Eng. Sci. Technol., Int. J.*, vol. 23, no. 2, pp. 264–273, Apr. 2020.
[3] L. Du, H. Hu, and Y. Wu, "Age factor removal network based on transfer learning and adversarial learning for cross-age face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 9, pp. 2830–2842, Sep. 2020.
[4] F. Horst, S. Lapuschkin, W. Samek, K.-R. Müller, and W. I. Schöllhorn, "Explaining the unique nature of individual gait patterns with deep learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, Dec. 2019.
[5] L. Yao, W. Kusakunniran, Q. Wu, J. Xu, and J. Zhang, "Collaborative feature learning for gait recognition under cloth changes," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Sep. 14, 2021, doi: 10.1109/TCSVT.2021.3112564.
[6] C. Benedek, B. Gálai, B. Nagy, and Z. Jankó, "Lidar-based gait analysis and activity recognition in a 4D surveillance system," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 1, pp. 101–113, Jan. 2018.
[7] U. Martinez-Hernandez, I. Mahmood, and A. A. Dehghani-Sanij, "Simultaneous Bayesian recognition of locomotion and gait phases with wearable sensors," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1282–1290, Feb. 2018.
[8] Y. Zhong, J. Wang, S. Wu, T. Jiang, Y. Huang, and Q. Wu, "Multilocation human activity recognition via MIMO-OFDM-based wireless networks: An IoT-inspired device-free sensing approach," *IEEE Internet Things J.*, vol. 8, no. 20, pp. 15148–15159, Oct. 2021.
[9] X. Ding, T. Jiang, Y. Zhong, Y. Huang, and Z. Li, "Wi-Fi-based location-independent human activity recognition via meta learning," *Sensors*, vol. 21, no. 8, p. 2654, Apr. 2021.
[10] M. J. Bocus, K. Chetty, and R. J. Piechocki, "UWB and WiFi systems as passive opportunistic activity sensing radars," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2021, pp. 1–6.
[11] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-Doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, Mar. 2006.
[12] H. Ji, C. Hou, Y. Yang, F. Fioranelli, and Y. Lang, "A one-class classification method for human gait authentication using micro-Doppler signatures," *IEEE Signal Process. Lett.*, vol. 28, pp. 2182–2186, 2021.

[13] S. Abdulatif, F. Aziz, K. Armanious, B. Kleiner, B. Yang, and U. Schneider, "Person identification and body mass index: A deep learning-based study on micro-Dopplers," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6.

[14] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 669–673, May 2018.

[15] Q. Chen, Y. Liu, F. Fioranelli, M. Ritchie, B. Tan, and K. Chetty, "DopNet: A deep convolutional neural network to recognize armed and unarmed human targets," *IEEE Sensors J.*, vol. 19, no. 11, pp. 4160–4172, Jun. 2019.

[16] R. Trommel, R. Harmanny, L. Cifola, and J. Driessen, "Multi-target human gait classification using deep convolutional neural networks on micro-Doppler spectrograms," in *Proc. Eur. Radar Conf. (EuRAD)*, 2016, pp. 81–84.

[17] S. Abdulatif, Q. Wei, F. Aziz, B. Kleiner, and U. Schneider, "Micro-Doppler based human-robot classification using ensemble and deep learning approaches," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2018, pp. 1043–1048.

[18] V. S. Papanastasiou, R. P. Trommel, R. I. A. Harmanny, and A. Yarovoy, "Deep learning-based identification of human gait by radar micro-Doppler measurements," in *Proc. 17th Eur. Radar Conf. (EuRAD)*, Jan. 2021, pp. 49–52.

[19] Z. Cao, Z. Li, X. Guo, and G. Wang, "Towards cross-environment human activity recognition based on radar without source data," *IEEE Trans. Veh. Technol.*, vol. 70, no. 11, pp. 11843–11854, Nov. 2021.

[20] G. Csurka, "A comprehensive survey on domain adaptation for visual applications," in *Domain Adaptation in Computer Vision Applications*. Cham, Switzerland: Springer, 2017, pp. 1–35.

[21] Y. Huang, Q. Wu, J. Xu, Y. Zhong, and Z. Zhang, "Unsupervised domain adaptation with background shift mitigating for person re-identification," *Int. J. Comput. Vis.*, vol. 129, no. 7, pp. 2244–2263, Jul. 2021.

[22] S. Cui, S. Wang, J. Zhuo, L. Li, Q. Huang, and Q. Tian, "Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3941–3950.

[23] Y. Zhu *et al.*, "Deep subdomain adaptation network for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2020.

[24] K. Zheng, C. Lan, W. Zeng, Z. Zhang, and Z.-J. Zha, "Exploiting sample uncertainty for domain adaptive person re-identification," 2020, *arXiv:2012.08733*.

[25] S. Li, M. Yuan, J. Chen, and Z. Hu, "AdaDC: Adaptive deep clustering for unsupervised domain adaptation in person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Oct. 5, 2021, doi: 10.1109/TCSVT.2021.3118060.

[26] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang, "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2507–2516.

[27] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2016.

[28] S. Cui, S. Wang, J. Zhuo, C. Su, Q. Huang, and Q. Tian, "Gradually vanishing bridge for adversarial domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12455–12464.

[29] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 443–450.

[30] H. Tang, K. Chen, and K. Jia, "Unsupervised domain adaptation via structurally regularized deep clustering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8725–8735.

[31] J. Goldberger, G. E. Hinton, S. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 17, 2004, pp. 513–520.

[32] 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.

[33] Y. Lang, Q. Wang, Y. Yang, C. Hou, H. Liu, and Y. He, "Joint motion classification and person identification via multitask learning for smart homes," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9596–9605, Dec. 2019.

[34] A. Jalalvand, B. Vandersmissen, W. De Neve, and E. Mannens, "Radar signal processing for human identification by means of reservoir computing networks," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6.

[35] S. Li, B. Xie, Q. Lin, C. H. Liu, G. Huang, and G. Wang, "Generalized domain conditioned adaptation network," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 1, 2021, doi: 10.1109/TPAMI.2021.3062644.

[36] F. Zhuang, X. Cheng, P. Luo, S. J. Pan, and Q. He, "Supervised representation learning: Transfer learning with deep autoencoders," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 4119–4125.

[37] J. Shen, Y. Qu, W. Zhang, and Y. Yu, "Wasserstein distance guided representation learning for domain adaptation," in *Proc. 32nd AAAI Conf. Artif. Intell. (AAAI)*, 2018, pp. 4058–4065.

[38] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 1180–1189.

[39] R. Xu, G. Li, J. Yang, and L. Lin, "Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1426–1435.

[40] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2066–2073.

[41] L. Zhang, J. Fu, S. Wang, D. Zhang, Z. Dong, and C. L. P. Chen, "Guide subspace learning for unsupervised domain adaptation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3374–3388, Sep. 2019.

[42] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 478–487.

[43] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

[44] K. G. Dizaji, A. Herandi, C. Deng, W. Cai, and H. Huang, "Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5736–5745.

[45] W. Hu, T. Miyato, S. Tokui, E. Matsumoto, and M. Sugiyama, "Learning discrete representations via information maximizing self-augmented training," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 1558–1567.

[46] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

[47] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[48] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The Mahalanobis distance," *Chemometrics Intell. Lab. Syst.*, vol. 50, no. 1, pp. 1–18, Jan. 2000.

[49] V. Koltchinskii, "Rademacher penalties and structural risk minimization," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1902–1914, Jul. 2001.

[50] Z. Yang and J. Laaksonen, "Regularized neighborhood component analysis," in *Proc. Scandin. Conf. Image Anal.*, 2007, pp. 253–262.

[51] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, 1967, vol. 1, no. 14, pp. 281–297.

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

[53] J. Liang, D. Hu, and J. Feng, "Domain adaptation with auxiliary target domain-oriented classifier," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16632–16642.

[54] J. Na, H. Jung, H. J. Chang, and W. Hwang, "FixBi: Bridging domain spaces for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1094–1103.

[55] S. Roy, E. Krivosheev, Z. Zhong, N. Sebe, and E. Ricci, "Curriculum graph co-teaching for multi-target domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5351–5360.

[56] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[57] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 8026–8037.

[58] D. Tan, K. Huang, S. Yu, and T. Tan, "Efficient night gait recognition based on template matching," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, vol. 3, 2006, pp. 1000–1003.

Yang Yang received the Ph.D. degree from Tianjin University, Tianjin, China, in 2019. Currently, he is an Assistant Professor with Tianjin University. His research interests include deep learning, pattern recognition, object detection, human activity recognition, and Micro-Doppler radar.

Xiaoyi Yang received the B.S. degree from the School of Computer Science, Central South University, China, in 2020. She is currently pursuing the M.A.Sc. degree with the School of Electrical and Information Engineering, Tianjin University, China. Her research interests include deep learning, person identification, and Micro-Doppler radar.

Takuya Sakamoto (Senior Member, IEEE) received the B.E. degree in electrical and electronic engineering from Kyoto University, Kyoto, Japan, in 2000, and the M.I. and Ph.D. degrees in communications and computer engineering from the Graduate School of Informatics, Kyoto University, in 2002 and 2005, respectively. From 2006 to 2015, he was an Assistant Professor with the Graduate School of Informatics, Kyoto University. From 2011 to 2013, he was also a Visiting Researcher with the Delft University of Technology, Delft, The Netherlands. From 2015 to 2019, he was an Associate Professor with the Graduate School of Engineering, University of Hyogo, Himeji, Japan. In 2017, he was also a Visiting Scholar with the University of Hawaii, Manoa, Honolulu, HI, USA. Since 2018, he has been a PRESTO Researcher with the Japan Science and Technology Agency, Kawaguchi, Japan. Currently, he is an Associate Professor with the Graduate School of Engineering, Kyoto University. His current research interests are systems theory, inverse problems, radar signal processing, radar imaging, and wireless sensing of vital signs. He was a recipient of the Best Paper Award from the International Symposium on Antennas and Propagation (ISAP) in 2012 and the Masao Horiba Award in 2016. In 2017, he was invited as a Semi-Plenary Speaker to the European Conference on Antennas and Propagation (EuCAP), Paris, France.

Francesco Fioranelli (Senior Member, IEEE) received the Ph.D. degree from Durham University, Durham, U.K., in 2014. He was a Research Associate with University College London from 2014 to 2016 and an Assistant Professor with the University of Glasgow from 2016 to 2019. He is currently a tenured Assistant Professor with TU Delft, The Netherlands. His research interests include the development of radar systems and automatic classification for human signatures analysis in healthcare and security, drones and UAVs detection and classification, automotive radar, wind farm, and sea clutter.

Beichen Li received the B.E. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2010, and the M.S. degree from the Stevens Institute of Technology (SIT), Hoboken, NJ, USA, in 2016. He is currently pursuing the Ph.D. degree in the field of deep learning for computer vision and the Internet of Things with Tianjin University, Tianjin, China. He was a Research Assistant with the Maritime Security Center, SIT, from 2014 to 2015. He has worked for several companies, including Good Technology, Tianjin; and Kingsoft Cloud, Beijing, China. His working experiences include quality assurance, software development, cloud service, data analysis, and big data processing.

Yue Lang received the Ph.D. degree from Tianjin University, Tianjin, China, in 2021. She was a Visiting Student at Nanyang Technological University, Singapore, from 2020 to 2021. She is currently a Lecturer with the Hebei University of Technology, Tianjin. Her research interests include Micro-Doppler effect, radar target recognition, and machine learning.