MCDCD: Multi-Source Unsupervised Domain Adaptation for Abnormal Human Gait Detection

Yao Guo, *Member, IEEE*, Xiao Gu, *Student Member, IEEE*, and Guang-Zhong Yang, *Fellow, IEEE*

***Abstract*—For gait analysis, especially for the detection of subtle gait abnormalities, the collected datasets involve high variability across subjects due to inherent biometric traits and movement behaviors, leading to limited detection accuracy and poor generalizability. To address this, we propose a novel deep multi-source Unsupervised Domain Adaptation (UDA) approach, namely Maximum Cross-Domain Classifier Discrepancy (MCDCD), which aims to improve the classification performance on the test subject (target domain) by leveraging the information from multiple labelled training subjects (source domains). Specifically, the proposed model consists of a feature extractor and a domain-specific category classifier per source domain. The former feature extractor learns to generate discriminative gait features. For the latter classifiers, we minimize the cross-entropy loss to accurately classify source samples, and simultaneously maximize a novel cross-domain discrepancy loss between any two category classifiers to minimize domain shift between multiple sources and the target domain. To validate the proposed MCDCD for detecting gait abnormalities on novel subjects, we collected both high-quality Motion capture (Mocap) and noisy Electromyography (EMG) data from eighteen subjects with both normal and imitated abnormal gaits. Experiment results using both data modalities demonstrate that the proposed approach can achieve superior performance in abnormal gait classification compared to baseline deep models and state-of-the-art UDA methods.**

***Index Terms*—Gait abnormality detection, multi-source unsupervised domain adaptation, cross-domain, mocap, EMG.**

I. Introduction

HUMAN gait abnormalities, mainly originating from the disorders of neurological or musculoskeletal systems, are prevalent among all population groups [1]. Among abnormal gait patterns, some commonly observed ones involving only subtle changes of the lower limb include the outward (supination) and inward (pronation) rotation of the ankle joint, as well as the outward (out-toeing) and inward (in-toeing) of the foot forward direction [2], [3]. Accurate detection of subtle gait changes, however, is challenging, and these changes are usually associated with disease manifestations and poor quality of life [4]. Hence, from a healthcare perspective, the early detection of subtle gait abnormality plays an essential role in numerous scenarios, ranging from elderly fall risk prediction, post-surgery functional assessment to rehabilitation [5].

Recent advances in deep learning technologies have enabled automatic extraction of discriminative features from a large amount of data [6]. Within this procedure, the inherent challenges are the lack of large-scale training datasets and, more importantly, subject-specific differences existing in both raw data and extracted features. Especially for the detection of abnormal gait patterns, the subtle changes indicating abnormalities could be less obvious than, if not interleaved by, the subject-specific patterns [7]. Hence, the classification model trained from small-scale datasets tends to either overfit on the training set or have the poor generalization capability to novel subjects, which limits its application in practice. This can be formulated as a problem in transfer learning [8], and without loss of generality, a subject with multiple collected gait samples can be regarded as a domain, and then the subject-specific differences can be regarded as the domain shift across domains (subjects). In order to increase the classification performance of the model on a novel subject (target domain), our objective is to mitigate the domain discrepancy between multiple sources and the unlabelled target domain.

To address aforementioned challenges, in this paper we propose a novel Maximum Cross-Domain Classifier Discrepancy (MCDCD) network to solve the challenging gait abnormality detection through multi-source Unsupervised Domain Adaptation (UDA) [9], [10]. The architecture of MCDCD consists of a feature extractor and domain-specific category classifiers corresponding to multiple sources as illustrated in Fig. 1. With a three-step adversarial training strategy, the feature extractor aims to generate common latent gait representations across domains and the domain-specific category classifiers aim to correctly classify the source samples. In the first step, the classification loss on source samples is minimized to update parameters of the feature extractor and category classifiers, and the triplet loss is incorporated to cluster samples within the same category yet from different sources more tightly. Following the idea as proposed in [10] and [11], the discrepancy loss on target samples is then maximized to update parameters of the category

Manuscript received November 28, 2020; revised March 29, 2021; accepted May 8, 2021. Date of publication May 14, 2021; date of current version October 5, 2021. This work was supported by the Science and Technology Commission of Shanghai Municipality under Grant 20DZ2220400 and in part by the Interdisciplinary Program of Shanghai Jiao Tong University under Grant YG2021QN117. *(Corresponding author: Guang-Zhong Yang.)*

Yao Guo and Guang-Zhong Yang are with the Institute of Medical Robotics and School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: yao.guo@sjtu.edu.cn; gzyang@sjtu.edu.cn).

Xiao Gu is with the Hamlyn Centre for Robotics Surgery, Imperial College London, SW7 2AZ London, U.K. (e-mail: xiao.gu17@imperial.ac.uk).

Digital Object Identifier 10.1109/JBHI.2021.3080502

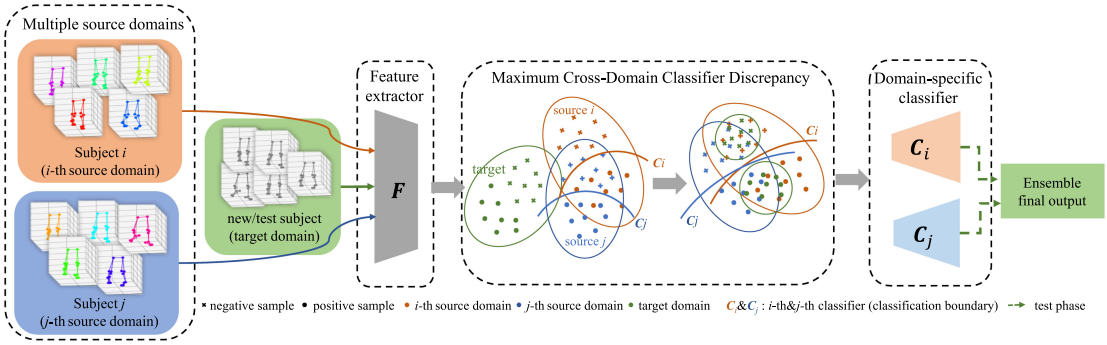


Fig. 1. An overview of the proposed framework for abnormal gait recognition. Here we demonstrate an example with two source domains (i.e., i-th and j-th domains). Our MCDCD network consists of a feature extractor and the domain-specific category classifiers. The feature extractor $\mathbf{F}$ maps the gait cycle data from all domains into discriminate feature vectors, and the category classifiers $\mathbf{C}_i\&\mathbf{C}_j$ aim to accurately classify the samples from i-th and j-th domains, respectively. The main idea of the proposed MCDCD network is to diminish the domain shift between multiple sources and the target domain by maximizing the cross-domain discrepancy loss, and meanwhile, to align the distributions of multiple sources by clustering the intra-class samples with the triplet loss.

classifiers, which aims to detect the unlabelled target samples that are far away from the support of the source domains. Instead of the discrepancy between a pair of classifiers for each source domain [10], [11], one of our contributions is to calculate the cross-domain discrepancy, thus enabling the alignment across multi-source distributions with a more compact architecture. In the final step, this cross-domain discrepancy loss is additionally minimized to update parameters of the feature extractor for generating target features that are close to the support of the source domains. By repeating these three steps, the parameters of the feature extractor and domain-specific category classifiers are gradually converged in an adversarial manner. In the test phase, the predicted label of a novel target sample is the ensemble output of all category classifiers.

To evaluate the proposed MCDCD, an abnormal gait dataset was constructed, in which five walking patterns (i.e., normal, supination, pronation, in-toeing, out-toeing) of eighteen subjects were simultaneously captured by Motion capture (Mocap) system and wearable Electromyography (EMG) sensors. Raw Mocap and EMG data were firstly segmented into gait cycles, followed by several preprocessing steps. Extensive experiments were conducted to demonstrate the superiority of the proposed MCDCD compared to various baselines and state-of-the-art UDA approaches.

In summary, the main contributions of this paper include:

- Leverage multi-source unsupervised domain adaptation to overcome the individual differences in the challenging abnormal gait detection.
- Propose a novel cross-domain discrepancy loss to facilitate the domain adaptation across multiple sources.
- Integrate the triplet loss into the multi-source UDA to cluster the intra-class samples, thus improving the recognition performance.

II. Related Works

A. Gait Analysis for Abnormality Detection

1) Gait Analysis Systems and Sensing Modalities: Most existing gait analysis systems capture the kinematic data of human walking styles by applying various sensing tools [12]. For clinical purposes, multi-camera Mocap systems are the gold standard solutions by accurately detecting 3D positions of reflective markers attached to the anatomical landmarks [5]. Because of the high cost and complicated settings, the Mocap system is usually limited to highly structured environments, such as laboratories or hospitals. To overcome this problem, rgb and/or depth cameras have been utilized to estimate 2D/3D skeletons directly from recorded images for gait analysis [13], [14], however, it is still challenging to apply such methods in clinical scenarios due to the unsatisfactory estimation of subtle lower limb movements [15]. Another line of gait analysis research by wearable sensors, such as inertial sensors, is enabled by attaching sensors to different body parts [16]–[19]; however, it typically requires sophisticated subject-dependent calibration before data collection.

Apart from kinematic data, surface EMG data, which measures the continuous contraction and relaxation of muscles from electrode-skin interfaces, provides a kinetics clue of human gait [20]. Although EMG signals have been widely applied for hand gesture [21] and gait phase recognition [22], [23], less attention has been paid on EMG-based abnormal gait recognition [24]. Besides, the force plates and pressure insoles have been used to perform gait analysis by measuring the ground reaction forces exerted to lower limbs [25]; however, they are difficult to infer the lower limb kinematics.

2) Conventional and Deep Gait Representation: In order to perform abnormal gait detection from the kinematic data, most conventional methods took advantage of features and gait models with obvious clinical clues [13], [14], [26]. In [13], the quantitative gait features {step length, gait cycle time, gait symmetric measure} were extracted. In addition, joint angle representation, which involves angles between any two body segments and/or angles with respect to the human body coordinate system, has been widely applied for abnormal gait recognition [14], [26]. They can provide an invariant and informative characterization of subtle gait changes. As raw EMG signals are commonly affected by several sources of noise, previous works explored various time-domain and frequency-domain features and selected those with high noise tolerance for different recognition tasks [21], [22], [24].

Although conventional methods can perform successfully in a specific clinical scenario, manually selecting the relevant features may lead to insufficient gait representation. With recent progress in Deep Neural Networks (DNN), automatically extracting deep representations has gained increasing popularity in the field of gait analysis. Recurrent Neural Networks (RNN) is advantageous in modeling sequential data with varied lengths. Among different RNN variations, Long Short-Term Memory (LSTM) is most popular for learning effective gait features [27]. Through the use of LSTM, Kidzinski *et al.* [28] performed gait event detection with joint angles and positions, and Luo *et al.* [23] took the raw EMG as input to achieve gait sub-phase recognition. Guo *et al.* [14] investigated the use of Bidirectional LSTM (BiLSTM) for recognizing abnormal gait patterns based on joint angle sequence input. In addition to RNN models, it has been well explored that Convolutional Neural Networks (CNN), especially for 1D-CNN that performs the convolution along the temporal axis, can achieve promising performance in processing biosignals [6]. In [29], the context-related gait features were directly translated from the raw inertial data with a deep CNN. Hu *et al.* [30] adopted the graph CNN (GCNN) to detect the freezing of gait, where GCNN can extract informative spatiotemporal information from joint trajectories. Besides, some recent works explored the combination of LSTM and CNN for improving the abnormal gait classification performance [31]. Despite the successful application of DNN models on gait analysis, simply training these models on human motion data would probably limit its generalization performance on a novel subject.

To resolve the generalization issue, conventional methods tend to engineer subject-invariant features as mentioned above [13], [14], [26], which however is limited by its insufficient representation. Horst *et al.* [32] successfully developed an explainable deep learning framework to explore the uniqueness of individual gait patterns in clinical biomechanics. Recently, Gu *et al.* [7] proposed a disentangled representation learning framework based on a multi-encoder autoencoder network to perform abnormal gait recognition, extracting the subject-invariant gait patterns through a cross-subject reconstruction training strategy. Nevertheless, the generalization capability of [7] would probably not hold without seeing any sample in the target subject; this issue can be potentially resolved by unsupervised domain adaptation introduced below.

B. Unsupervised Domain Adaptation

One of the inherent challenges of deep learning is that the performance of the learned model tends to significantly degrade while testing on novel domains/subjects with the existence of domain shift. Domain Adaptation (DA) aims to diminish the domain shift between the training (source) and test (target) domains by matching their representations [33]. Among these, unsupervised domain adaptation tackles a more challenging scenario where all available target samples are unlabelled [34]. Most of the previous methods focus on solving the UDA with a single source domain, and they have achieved promising results in different computer vision tasks. However, training data are in practice collected from multiple sources/subjects, especially in healthcare tasks. Recently, studies on multi-source UDA have gained increasing popularity [9]–[11], [35].

1) Single-Source UDA: The most common way for UDA is to eliminate the domain shift between source and target domains through the metric learning mechanisms. Some works achieved UDA by minimizing the Kullback-Leibler divergence, $\mathcal{H}$ divergence [33], as well as Maximum Mean Discrepancy (MMD) measuring the divergence in high-dimensional space [36], [37]. Another popular category is adversarial-based approaches. Domain Adversarial Neural Network (DANN) [38] introduced a domain classifier to discriminate the latent feature from source to target, where the feature extractor was trained to generate the latent features that can be correctly classified by category classifier yet can simultaneously deceive the domain classifier. The adversarial training can be easily realized by a gradient reversal layer. In addition, Saito *et al.* proposed the Maximum Classifier Discrepancy (MCD) for UDA, aligning the distributions of source and target by considering task-specific decision boundaries between classes [11]. The MCD consists of a feature extractor and a pair of category classifiers. During the training, the discrepancy between two classifier outputs is maximized to detect the target samples excluded by the support of the source, while the feature extractor is optimized by minimizing the discrepancy for generating target features that are close to the support. Other commonly used frameworks are based on the Generative Adversarial Networks (GANs), which are able to generate fake data from target domains as augmented low-level representations [39], [40].

2) Multi-Source Unsupervised Domain Adaptation (MS-UDA): Increasing attention has been posed on MS-UDA to achieve the alignment between multiple source domains and a single target domain [9], [33]. Earlier study [33] provided a theoretical investigation on such attempt by introducing the $\mathcal{H}\triangle\mathcal{H}$-distance between the weighted combination of source domains and target domain. Directly extended from single-source UDA, one line of research performed the UDA between each source domain and the target domain respectively and subsequently chose the model with the best performance. It may also involve simply performing UDA between the data combined from all source domains and that from the target domain. However, these methods ignore the semantic relationship between source domains and cannot leverage these sources effectively. More advanced, Zhao *et al.* [35] proposed a Multi-source Domain Adversarial Network (MDAN) based on adversarial training strategy, which is the extension of the popular DANN to multiple source domains. Deep CockTail Network (DCTN) [41] incorporated a multi-way adversarial training and multi-source category classifier to battle the domain and category shifts among multiple sources. More recently, Peng *et al.* [10] proposed a Moment Matching for Multi-Source Domain Adaptation (M^3SDA) approach by extending the popular MCD to a multi-source version and introduced the moment matching distance to transfer knowledge learned from multiple sources to the target domain. Although MS-UDA approaches have achieved promising results in computer vision, their capability and feasibility have not been well explored in handling the significant inter-subject heterogeneity existing in biosignals.

III. Methodology

A. Problem Statement

Considering that we have N labelled source domains (training subjects) $\mathcal{D}_S = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N\}$ and an unlabelled target domain (test subject) $\mathcal{D}_T$, the main objective is to mitigate the domain shift between unlabelled target samples $\mathbf{X}_T$ collected from $\mathcal{D}_T$ and those samples $\mathbf{X}_S = \{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ collected from multiple source domains $\mathcal{D}_S$, thus improving the accuracy of the final prediction. $\mathbf{Y}_S = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_N\}$ indicates the corresponding ground truth labels of $\mathbf{X}_S$.

B. Maximum Cross-Domain Classifier Discrepancy

The overall framework of the proposed MCDCD is presented in Fig. 1. The feature extractor $\mathbf{F}$ maps the samples from both $\mathcal{D}_S$ and $\mathcal{D}_T$ into latent feature vectors, which are the input into the classifier. Different from previous methods like MCD [11] and M ^{3} SDA [10], using a pair of classifiers for each source domain, only one classifier $\mathbf{C}_i$ is constructed for i-th source domain in our network, where $i = 1, \ldots, N$ and N is the number of source domains (training subjects). It is noted that the proposed framework is flexible in choosing different networks as $\mathbf{F}$ and $\mathbf{C}$. Based on the proposed model, we follow a three-step training procedure as proposed in [10], [11], which can not only classify source samples correctly and align the target domain with respect to multiple source domains simultaneously, but also reduce the training memory compared to those using paired classifiers. The overall training scheme is to repeat these three steps until convergence.

1) Step A - Minimize Classification Loss on Source: In the first step, the feature extractor $\mathbf{F}$ and domain-specific category classifiers $\{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ are trained simultaneously with respect to the source samples $\{(\mathbf{X}_S, \mathbf{Y}_S)\}$. Accordingly, $\mathbf{F}$ can generate the discriminative features related to the task, and each classifier can classify the source samples correctly. The minimization objective is as follows:

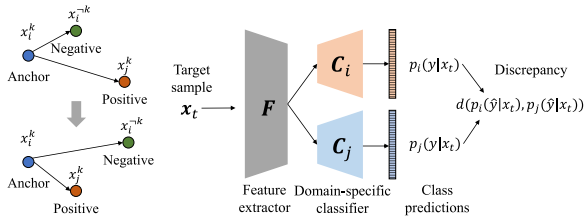
$$\min_{\mathbf{F},\{\mathbf{C}_i\}} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} [\mathcal{L}_{\mathbf{C}_i}(\mathbf{X}_j, \mathbf{Y}_j) + \mathcal{L}_{trip}(\mathbf{X}_i, \mathbf{X}_j)] \quad (1)$$

where $\mathcal{L}_{trip}(\mathbf{X}_i, \mathbf{X}_j)$ is the triplet loss between the samples from i-th and j-th domains. $\mathcal{L}_{\mathbf{C}_i}(\mathbf{X}_j, \mathbf{Y}_j)$ is the classification loss of $\mathbf{C}_i$ with respect to source samples from j-th domain and in this paper we choose the cross-entropy loss as follows:

$$\mathcal{L}_{\mathbf{C}_i}(\mathbf{X}_j, \mathbf{Y}_j) = -\mathbb{E}_{(x_j, y_j)\sim(\mathbf{X}_j, \mathbf{Y}_j)} \sum_{k=1}^{K} \mathbf{1}_{[k=y_j]} \log p_j(y|x_j) \quad (2)$$

where K is the number of classes to be recognized. In specific, we calculate the cross-domain classification loss, which is aimed at training each domain-specific category classifier to correctly classify the samples from not only its own domain but also other sources.

The triplet loss has been proven to be beneficial for training the samples with small differences [42]. As shown in Fig. 2(a), it minimizes the distance between the anchor and the positive sample and simultaneously maximizes the distance from the anchor to the negative sample. Especially for abnormal gait recognition, the subtle gait abnormality could be less obvious than the subject differences. Therefore, the triplet loss $\mathcal{L}_{trip}$ is adopted to cluster the gait features with the same label more tightly, which can be formulated as follows:



(a) Triplet loss (b) Cross-domain discrepancy loss

Fig. 2. Illustration of the cross-domain discrepancy loss and the triplet loss. (a) Triplet loss is aimed at clustering the intra-class samples across domains. (b) Cross-domain discrepancy loss calculates the discrepancy between the probabilistic outputs of two domain-specific category classifier with the input of a target sample x_t.

$$\mathcal{L}_{trip}(\mathbf{X}_i, \mathbf{X}_j) = \sum_{k=1}^{K} \mathbb{E}_{(x_i, x_j)\sim(\mathbf{X}_i, \mathbf{X}_j)} [||\mathbf{F}(x_i^k) - \mathbf{F}(x_j^k)|| - ||\mathbf{F}(x_i^k) - \mathbf{F}(x_i^{\neg k})|| + \gamma] \quad (3)$$

where the superscript $(k, \neg k)$ indicates the class index of a sample x, and γ is the margin constraint. Through the use of the triplet loss, intra-class samples from different source domains will have smaller distance ($\leq \gamma$) than those inter-class samples, which could help to learn subject-invariant features and to facilitate the alignment across multiple domains.

2) Step B - Maximize Discrepancy Loss on Target: In this step, we aim to only train the domain-specific category classifiers $\{\mathbf{C}_i\}$ with the unlabelled target samples while fixing the parameters of the feature extractor $\mathbf{F}$. Previous works calculated the discrepancy between the paired classifier outputs for each source domain, and they found that the maximization of the discrepancy can help to detect the target samples excluded by the source boundaries [10], [11]. Differently, we introduce the calculation of the cross-domain discrepancy d_{dis} indicating the discrepancy between the outputs of any two domain-specific category classifiers; this strategy can be viewed as another contribution of this paper. Moreover, the cross-domain discrepancy enables a better alignment across multiple sources.

Fig. 2(b) demonstrates an example for calculating the discrepancy between i-th and j-th source domains. Given domain-specific category classifiers $\mathbf{C}_i$ and $\mathbf{C}_j$ for i-th and j-th source domains and a target sample x_t, the classifiers' probabilistic outputs are $p_i(\hat{y}|x_t)$ and $p_j(\hat{y}|x_t)$, respectively. Then the discrepancy loss can be defined as the absolute value of the difference between two outputs:

$$d_{dis}(p_i, p_j) = \frac{1}{K} \sum_{k=1}^{K} |p_i^k(\hat{y}|x_t) - p_j^k(\hat{y}|x_t)|$$

$$\mathcal{L}_{dis}(\mathbf{X}_T) = \mathbb{E}_{x_t\sim\mathbf{X}_T} \left[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} d_{dis}(p_i, p_j) \right] \quad (4)$$

Previous studies have proven that the overall performance drops significantly without adding the classification loss in this step [10], [11]. Inspired by this, the optimization objective can be written as follows:

$$\min_{\{\mathbf{C}_i\}} \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_{\mathbf{C}_i}(\mathbf{X}_i,\mathbf{Y}_i) - \mathcal{L}_{dis}(\mathbf{X}_T) \tag{5}$$

where the maximization of discrepancy is converted to a minimization problem by taking its negative value. Notably, to accelerate the training process, we calculate the classification loss of the classifiers with samples from their own domains.

3) Step C - Minimize Discrepancy Loss on Target: The last step aims to train the feature extractor $\mathbf{F}$ to minimize the discrepancy while the parameters of domain-specific category classifiers are fixed. In order to balance the tradeoff between the convergence speed of feature extractor and category classifiers in an adversarial manner, we repeated this step with n_{iter} times with the same minibatch data. The objective function is:

$$\min_{\mathbf{F}} \mathcal{L}_{dis}(\mathbf{X}_t) \tag{6}$$

Notably, the adversarial training of feature extractor and category classifiers is performed with alternative updates of these three steps.

C. Ensemble Output

In the test phase, the parameters of the feature extractor $\mathbf{F}$ and domain-specific category classifiers $\{\mathbf{C}_i\}$ are fixed. As there are N category classifiers, the final prediction on the target samples are derived through an ensemble strategy, which takes the weighted average of the probabilistic outputs:

$$p(y|\mathbf{X}_t) = \sum_{i=1}^{N}\omega_i p_i(y|\mathbf{X}_t), \sum_{i=1}^{N}\omega_i = 1 \tag{7}$$

where p_i is the probability provided by classifier $\mathbf{C}_i$ and $\omega_i = p_i(y|\mathbf{X}_i)/\sum_{j=1}^{N} p_i(y|\mathbf{X}_j)$ is the weight. Then the overall accuracy of MCDCD can be derived from $p(y|\mathbf{X}_t)$.

D. Theoretical Insights

1) Single-Source MCD: Our approach is expanded from MCD [11], which was proposed to perform adversarial training for single-source unsupervised domain adaptation. Different from those approaches based on additional domain classifiers [34], MCD utilized two category classifiers expected to achieve similar predictions on the target domain.

Based on the bounding condition theorem proposed by Ben-David *et al.* [33], the target error ϵ_T of a classifier h can be upper bounded by the sum of three terms. They are the source error ϵ_S of h, the distance in the symmetric difference hypothesis space $\mathcal{H}$ between two classifiers $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S,\mathcal{D}_T)$, and the combined error of the ideal joint classifier on two domains λ:

$$\epsilon_T(h) \le \epsilon_S(h) + \frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S,\mathcal{D}_T) + \lambda \tag{8}$$

The focus is to constrain $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S,\mathcal{D}_T)$, which is the upper bound of two classifiers belonging to the hypothesis space $\mathcal{H}$. It is defined in [11] as below,

$$d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S,\mathcal{D}_T) = 2\sup_{(h,h')\in\mathcal{H}^2}\Big|\mathop{\mathbb{E}}_{x\sim S} I[h(x)\neq h'(x)] - \mathop{\mathbb{E}}_{x\sim T} I[h(x)\neq h'(x)]\Big| \tag{9}$$

As the source domain is trained with labelled data, the term $\mathbb{E}_{x\sim S}\, I[h(x)\neq h'(x)]$ is is assumed to be very low, and then Eq. (9) can be approximated as $2\sup \mathop{\mathbb{E}}_{x\sim T} I[h(x)\neq h'(x)]$, which means the disagreement of two classifiers on the target domain. As such, the objective is to train two classifiers which achieve agreement in the target domain, whereas $\mathbb{E}_{x\sim S}\, I[h(x)\neq h'(x)]$ and $\mathbb{E}_{x\sim T}\, I[h(x)\neq h'(x)]$ is as close as possible. In practice, MCD [11] applied a feature generator G and paired classifiers C_1, C_2 to minimize the supremum. G & C_1 can be viewed as h, whereas G & C_2 as h'. To minimize the supermum can be considered as the min-max optimization, $\min_G \max_{C_1,C_2} \mathbb{E}_{x\sim T}\, I[G\circ C_1(x)\neq G\circ C_2(x)]$. Such min-max optimization is realized by the iterative adversarial training strategy.

2) MCDCD: For multi-source UDA, there are N labelled source domains $\mathcal{D}_S = \{\mathcal{D}_1,\mathcal{D}_2,\ldots,\mathcal{D}_N\}$. For each $i\in\{1,\ldots,N\}$, let S_i be the labelled sample set from source $\mathcal{D}_i$. According to [33], the $d_{\mathcal{H}\Delta\mathcal{H}}$ bound for multi-source UDA without considering the constant term can be formulated as below,

$$\epsilon_T(h) \le \sum_{i=1}^{N}\alpha_i(\epsilon_i(h) + \frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_i,\mathcal{D}_T)) + \lambda_\alpha \tag{10}$$

$$\begin{aligned}
&\sum_{i=1}^{N} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_i,\mathcal{D}_T)\\
&= 2\sum_{i=1}^{N}\sup_{(h_i,h_i')\in\mathcal{H}^2}\Big|\mathop{\mathbb{E}}_{x\sim S_i} I[h_i(x)\neq h_i'(x)] - \mathop{\mathbb{E}}_{x\sim T} I[h_i(x)\neq h_i'(x)]\Big|\\
&\le \sum_{(i,j)\in N^2}\sup_{(h_i,h_j)\in\mathcal{H}^2}\Big|\mathop{\mathbb{E}}_{x\sim S_i} I[h_i(x)\neq h_j(x)] - \mathop{\mathbb{E}}_{x\sim T} I[h_i(x)\neq h_j(x)]\Big|
\end{aligned} \tag{11}$$

h_i can be viewed as the domain specific category classifier for source domain $\mathcal{D}_i$. Similarly, $\mathbb{E}_{x\sim S_i}\, I[h_i(x)\neq h_j(x)]$ can be minimized by supervised cross-domain loss between paired source domains $\mathcal{D}_i$ and $\mathcal{D}_j$. In the implementation, it is realized by Eq. (1). Therefore, the objective can be approximated as to minimize the supremum of $\sum_{(i,j)\in N^2}|\,\mathbb{E}_{x\sim T}\, I[h_i(x)\neq h_j(x)]|$. Overall, to minimize such supremum is equivalent to the min-max optimization realized by an iterative adversarial training between the feature extractor $\mathbf{F}$

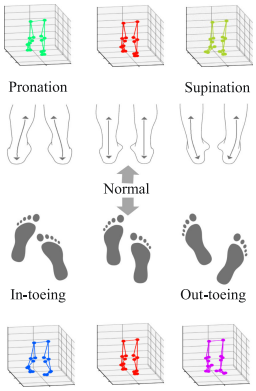


Fig. 3. Illustration and visualization of the five categories of gait patterns in the self-collected gait dataset, including the normal walking style and four abnormal patterns (i.e., in-toeing, out-toeing, supination, and pronation).

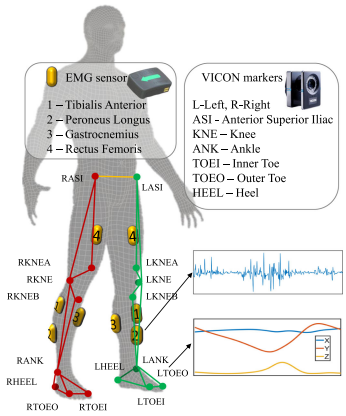


Fig. 4. Demonstration of sensor displacement. 8 EMG sensors were attached on the central position of muscles {Tibialis Anterior, Peroneus Longus, Gastrocnemius, and Rectus Femoris}. 16 markers were attached to the anatomical landmarks for being captured by the Mocap system.

and category classifiers $\{\mathbf{C}_i\}$ for the loss $\mathcal{L}_{dis}$, similar to the single-source version.

IV. Abnormal Gait Dataset and Data Processing

A. Data Collection

Eighteen healthy subjects (two females and sixteen males) without any known lower limb disabilities or recent injuries were recruited, with informed consent given before participating in the experiment. This gait experiment was approved by Imperial College Research Ethics Committee under Reference ICREC-18IC4915.

As illustrated in Fig. 3, each subject was asked to walk naturally and imitate four abnormal gait patterns (**in-toeing**, **out-toeing**, **supination**, and **pronation**) with the normal speed [14], where the commercial corrective insoles were used to help the imitation of supination and pronation more naturally. The participants walked at a natural speed and no additional requirements on the gait speed were imposed on the participants. During the data collection, we simultaneously collected both Mocap (Vicon Motion System Ltd., Oxford, U.K.) and EMG (Delsys Trigno Avanti wireless EMG system, Delsys Incorporated, USA) data. Fig. 4 shows the displacement of 16 reflective markers on the anatomical landmarks of the lower limb and eight EMG sensors for measuring the muscle activities. The Mocap and EMG systems were well-synchronized through Vicon Locklab, where the sampling rates are $120Hz$ and $1200Hz$, respectively.

Every subject repeated eight trials per gait pattern, and in each trial they walked back and forth along the diagonal line of a $3m \times 3m$ area. Note that the EMG data of subjects 6 & 17 were not collected due to personal reasons. In total, we collected $N_{Mocap} = 720 = 18(subjects) \times 5(patterns) \times 8(trials)$ and $N_{EMG} = 640$ in our abnormal gait dataset.

B. Mocap Data Processing

The Mocap system captures the high-quality 3D positions $\mathbf{p}(t) = [x, y, z]^{\mathrm{T}}$ of the attached markers at each frame t. This is also known as the 3D skeleton representation [43]. For a walking sample, the Mocap data can be defined as $\mathbf{x}_{mocap} = [\mathbf{p}_1^{\mathrm{T}}, \ldots, \mathbf{p}_J^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{3J \times M}$, where J is the number of markers and M is the number of frames.

1) Joint Angles: Joint angles have been widely used for representing gait movement [14], [26] due to its invariance to rotation and translation of the raw data. Following [14], a human body coordinate system was first built based on {LASI, RASI} joints and the instantiate velocity vector, which can further determine the Sagittal, Coronal, and Transverse planes of body. Accordingly, 30 joint angles were calculated, including the angles between any two links and the angles between the link to the normal vector of each plane. The final joint angle representation can be expressed as $\mathbf{x}_{angle} \in \mathbb{R}^{30 \times M}$.

2) Joint Trajectories: In addition to the angle representation, joint trajectories captured from Mocap system can also be used as the gait representation [7]. However, raw trajectories are not invariant to the rotation and translation in 3D space, which means that several preprocessing steps are needed before feeding them into deep models. We first rotated the human body coordinate system to a identity matrix and moved the center to the zero point, thus removing the rotation and translation variances. Thus, the normalized joint trajectory representation is $\mathbf{x}_{trj} \in \mathbb{R}^{48 \times M}$.

C. EMG Data Processing

The raw EMG data representing a walking sample can be denoted as $x_{raw} \in \mathbb{R}^{K \times M}$, where M is the total frames of the EMG sequences and K indicates the channels of EMG sensors ($K = 8$ in this paper). The detrending step was performed to remove the motion artifacts from the raw EMG data.

1) Envelope of EMG Data: As raw EMG signals are typically contaminated with various kinds of noise, the smooth envelope x_{env} was also extracted to filter the noise out from the raw data. It is achieved by using the root mean square (RMS) filter with a time window of $20ms$. The EMG envelope sequences have the same dimensions as the raw data.

2) Time-Domain and Frequency-Domain Features: Except for the sequence representation of the EMG data, in the past

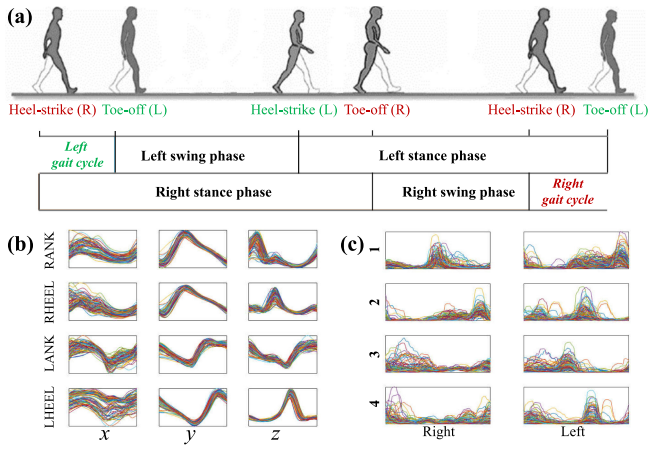


Fig. 5. Illustration of the gait cycle and abnormal gait data within a cycle. (a) A gait cycle that can be determined by two consecutive events of the same foot. In this paper, we detected right toe-off events from Mocap data and segmented both Mocap and EMG data; (b) Exemplar trajectories of four joints (RANK, RHEEL, LANK, LHEEL) within a gait cycle across various patterns&subjects captured from Mocap system; (c) EMG envelopes of eight channels within a gait cycle in terms of all patterns&subjects.

decades, both time-domain and frequency-domain features were widely used for EMG-based motion analysis, which have been proven to be robust to noise [21], [44]. To this end, we followed [21], [44] to calculate twenty time&frequency domain meta-features $x_{meta} \in \mathbb{R}^{20\times1}$, which are presented in Table S-A in the supplementary materials.

D. Gait Cycle Representation

The human walking data can be represented by the repetitive movements - *gait cycle*. A gait cycle indicates the data between two successive gait events (e.g., toe-off or heel-strike), which can be further divided into two phases based on the foot contact with the ground as shown in Fig. 5(a): stance phase and swing phase. In this paper, we segment all gait data ($\mathbf{x}_{angle}, \mathbf{x}_{trj}, \mathbf{x}_{raw}$, and $\mathbf{x}_{env}$) into gait cycle segments by detecting the right toe-off events from the Mocap data. Figs. 5(b) and (c) demonstrate the examples of $\mathbf{x}_{trj}$ and $\mathbf{x}_{env}$ within a cycle. Noted that those cycles containing turn-around or gait initiation/stop events were excluded. Recall that the gait data are temporal sequences (varied M across samples), we then used interpolation and resampling to reshape the length of the gait data into $M' = 128$, which help to the capability of different models for extracting latent features. Hence, the final gait representations are $\mathbf{x}_{angle} \in \mathbb{R}^{30\times128}$, $\mathbf{x}_{trj} \in \mathbb{R}^{48\times128}$, ($\mathbf{x}_{raw}$, $\mathbf{x}_{env}$) $\in \mathbb{R}^{8\times128}$. Subsequently, the z-score normalization was then performed per channel of Mocap and EMG data. Further details on gait cycles per pattern/subject in the dataset are available from the supplementary materials.

V. Experiments

A. Implementation Details

1) Experimental Settings: The experiments were performed on a desktop with Intel(R) Core(TM) i7-7700 K CPU4.2 GHz and an NVIDIA Titan 1080ti GPU. The proposed MCDCD method and the compared methods were implemented by PyTorch. In specific, we adopted the adaptive momentum (Adam) for optimizing the model parameters, where the initial learning rate $lr = 0.001$ and $\beta = \{0.9, 0.999\}$.

2) Network Architecture: In this paper, we compared both CNN and RNN as the feature extractor for generating discriminative gait features. The input of the feature extractor $\mathbf{F}$ is the final gait representations as mentioned in Section IV-D.

For RNN, we evaluated BiLSTM and BiLSTM with attention layers (BiLSTM_attn). Specifically, an architecture with two BiLSTM layers was used. We set the number of hidden units $N_h = \{512256\}$ and the dropout rate between two BiLSTM layers $r_{dp} = 0.2$.

For CNN, we adopted 1D-CNN as the feature extractor $\mathbf{F}$ to perform the convolution operation along the temporal axis. The architecture of $\mathbf{F}$ consists of two convolution (*conv*) layers and is presented as (*input*, *output*, *kernel*, *stride*, *padding*). The two *conv* layers are: *conv1* - (d_{in}, 64, 8, 1, 0) and *conv2* - (64, 128, 8, 1, 0), where d_{in} is the dimension of the gait data as introduced in Section IV-D. Following each *conv* layer, there are two standard layers (i.e., batch normalization layer and the ReLU layer).

The final output of the feature extractor was then reshaped into a 1D latent feature vector as the input into the category classifier. We simply utilized two fully connected (*fc*) layers - (*fc1*, *bn*, *relu*, *fc2*) as the classifier, where we set *fc1* (v_{in}, 128) and *fc2* (128, 5) and the dropout rate of between two *fc* layers $r_{dp} = 0.2$. For the comparison methods (DANN [38] and MDAN [35]), the similar network with two *fc* layers was used as the domain classifier and *fc2* was set to (128, 2).

3) Evaluation Protocol and Metrics: To provide valuable insight into the generalization capability of the models applied to a novel subject, the leave-one-subject-out (LOSO) cross-validation protocol was adopted. Given N subjects in the dataset, we held one subject out as the test subject in each run and repeated the recognition experiments N runs. The gait data from the remaining $N-1$ subjects form the source domains, and those from the held-out one as the target domain. Especially for the test subject (target domain), we took turns to use 50% data to perform UDA training and the other 50% as the final test set. In terms of the baseline methods without the use of DA, we also split one training subject out as the validation set.

In terms of metrics, we used the recognition accuracy $\texttt{Acc}$=(TP+TN)/(TP+TN+FP+FN)$\times100\%$, to provide an overall evaluation on the recognition performance, where TP, TN, FP and FN are the numbers of true positive, true negative, false positive, and false negative, respectively. Moreover, for the performance on each gait pattern, we also calculated Precision $\texttt{prec}$=TP/(TP+FP), Recall $\texttt{rec}$=TP/(TP+FN), and F1 score. The final reported results are the averaged value over five times of the LOSO cross-validation experiments, where in each time the random seed and split were used.

B. Comparison Results on Feature Extraction Methods

First, we conducted the comparison of various deep feature extraction methods in terms of different gait data representations for abnormal gait recognition. The results are listed in

TABLE I
COMPARISON OF FEATURE EXTRACTION METHODS

	Features	Classifier	TrainAcc	ValAcc	TestAcc
Mocap	hist. of joint angles	SVM	98.67	80.97	82.45
	joint angles	BiLSTM	99.84	83.21	82.73
	joint angles	BiLSTM_attn	99.82	85.9	84.56
	joint angles	1D-CNN	99.84	88.89	88.96
	joint trajectory	BiLSTM	99.66	88.33	88.34
	joint trajectory	BiLSTM_attn	99.62	89.89	89.08
	joint trajectory	1D-CNN	99.87	**92.86**	**91.96**
EMG	meta-features	SVM	89.45	35.25	34.18
	raw data	BiLSTM	79.40	34.62	35.47
	raw data	BiLSTM_attn	77.61	35.56	36.73
	raw data	1D-CNN	97.94	39.61	39.02
	envelope of raw data	BiLSTM	97.06	36.65	37.86
	envelope of raw data	BiLSTM_attn	94.81	39.78	39.84
	envelope of raw data	1D-CNN	99.00	**42.06**	**42.64**

Table I. In specific, we evaluated the performance of **1D-CNN**, **BiLSTM**, and **BiLSTM_attn** as the feature extractor to learn the discriminative feature from either Mocap or EMG data. To ensure a fair comparison, the same category classifier was adopted by directly taking the latent feature vector as the input. Besides, two baseline methods using hand-crafted features for gait abnormality detection were also compared, where the histogram of joint angles was extracted from Mocap data [14] and time-&frequency- domain meta-features was calculated from raw EMG data [21], [22]. Following our previous work, the Support Vector Machine (SVM) was chosen for classification.

As shown in Table I, deep feature extractors all outperform the conventional baseline methods, demonstrating the superiority of the deep models as the feature extractor. In addition, we can also observe that **1D-CNN** outperforms the **BiLSTM** and **BiLSTM_attn**. For Mocap data, two gait data representations (i.e., joint angles and joint trajectory) are compared, where the recognition results of using joint trajectories are better than those of using joint angles (**1D-CNN**: trajectory - 91.96%; angles - 88.96%). Although joint angle representation is invariant to the translational and rotational variations of the human body [14], compared to the normalized joint trajectory representation, it may fail to capture those subtle changes and lead to less discriminative descriptions. On the other hand, the EMG envelope representation achieves better recognition performance compared to the noisy raw EMG data. Accordingly, in the following, we only report the results using the joint trajectory representation and the envelope of EMG.

C. Comparison Results on State-of-The-Art Approaches

Next, to demonstrate the effectiveness of the proposed MCDCD for recognizing abnormal gait patterns on a novel subject, a detailed comparison with respect to several baselines, state-of-the-art UDA approaches, and a domain generalization method was conducted. To ensure a fair comparison, all the compared methods based on UDA were implemented with the same network architecture.

- Baseline methods: The architecture of the baseline methods consists of a feature extractor and a category classifier. Here, three baselines were compared by using **1D-CNN**, **BiLSTM**, and **BiLSTM_attn** (BiLSTM with the attention layer) as the feature extractor. For other compared methods, we chose 1D-CNN as **F**.
- **DANN** - Domain Adversarial Neural Network [38]: A single-source UDA method. In addition to the basic network structure, DANN contains another domain classifier to discriminate whether a sample belongs to the source domain or target domain.
- **MCD** - Maximum Classifier Discrepancy [11]: A single-source UDA method. The architecture of MCD consists of a feature classifier and two category classifiers, which is aimed at maximizing the discrepancy between two classifier outputs of the target samples that are far from the support of the source.
- **MDAN** - Multi-source Domain Adversarial Network [35]: An extended version of DANN to the multi-source scenario, which includes N domain classifiers (N is the number of source domains - training subjects).
- **MS-MCD** - Multi-Source Maximum Classifier Discrepancy: The multi-source version of the standard MCD [11], where the paired category classifiers are adopted for each source domain. Noted that **MS-MCD** can also be regarded as an ablated model of our proposed **MCDCD** that without $\mathcal{L}_{trip}$ and with paired category classifiers.
- **M ^{3} SDA** - Moment Matching for Multi-Source Domain Adaptation [10]. In addition to **MS-MCD**, they introduced the moment matching loss to better leverage multi-source information.
- **Disentangle** [7]. On top of UDA methods, a state-of-the-art method for generalizable abnormal gait recognition, which disentangles the subject-specific feature from abnormal pattern-specific gait features, was also compared. It applied a multi-encoder autoencoder architecture and fulfilled the disentanglement by cross-subject reconstruction. This method was evaluated on the same abnormal gait dataset. As its proposed cross-modal transfer, which utilizes Mocap data to learn clean representations, is out of the scope of this paper, we only compared the results using the single-modal network.

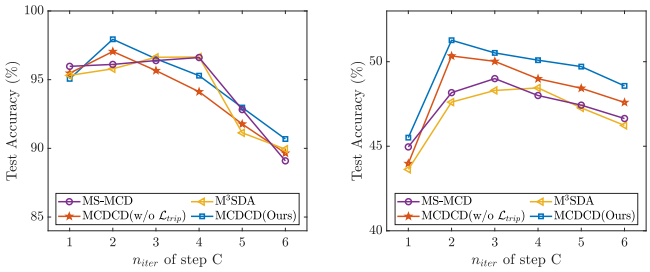
Following previous works [10], [41], two evaluation standards were used for single-source UDA methods (i.e., DANN and MCD): 1) *single best*, reporting the single best source transfer result on the test set; 2) *source combine*, combining multiple sources into a single domain. In addition to the comparison with state-of-the-art methods, we also evaluated the performance of an ablated version (without the use of triplet loss) of the proposed approach, referred to MCDCD (w/o $\mathcal{L}_{trip}$).

1) Evaluation of n_{iter}: Notably, there is a hyperparameter n_{iter} in the proposed MCDCD as well as the MS-UDA methods extended from MCD (i.e., MS-MCD and M ^{3} SDA), which indicates the repeat times of minimizing the discrepancy loss on target samples. As shown in Fig. 6, we first evaluated $n_{iter} \in \{1, 2, 3, 4, 5, 6\}$ of MS-MCD, M ^{3} SDA, and our MCDCD so as to choose the optimal one for the comparison experiments. For both Mocap and EMG data, our MCDCD and its ablated network achieve the best recognition rates when $n_{iter} = 2$, while MS-MCD and M ^{3} SDA perform best at $n_{iter} = 3$ or 4. This

TABLE II
COMPARISON RESULTS ON ABNORMAL GAIT RECOGNITION USING MOCAP & EMG DATA

Data	Category	Methods	Normal			Supination			Pronation			In-toeing			Out-toeing			TestAcc(%)
			prec	rec	F1	prec	rec	F1	prec	rec	F1	prec	rec	F1	prec	rec	F1	
Mocap	Baselines	BiLSTM	.803	.837	.819	.837	.805	.821	.873	.865	.869	.954	.955	.955	.956	.969	.964	88.84
		BiLSTM_attn	.814	.838	.826	.834	.836	.835	.892	.849	.870	.955	.958	.957	.946	.962	.954	89.08
		1D-CNN	.867	.906	.886	.902	.837	.868	.903	.913	.908	.948	.968	.958	.967	.970	.969	91.96
	SS-UDA	DANN [38]†	.774	.847	.808	.849	.815	.832	.876	.799	.836	.894	.922	.908	.881	.898	.890	85.64
		DANN [38]$^{\#}$	.855	.907	.880	.888	.841	.864	.899	.892	.896	.972	.971	.971	.953	.964	.959	91.54
		MCD [11]†	.764	.846	.803	.898	.878	.888	.868	.881	.875	.874	.849	.862	.869	.925	.897	86.61
		MCD [11]$^{\#}$	.910	.893	.901	.919	.831	.873	.915	.873	.893	.919	.983	.950	.895	.972	.932	91.07
	MS-UDA	MDAN [35]	.908	.928	.918	.936	.896	.915	.928	.936	.932	.977	.979	.978	.959	.971	.965	94.23
		MS-MCD	**.956**	.920	.938	.944	.932	.938	.956	.929	.942	.980	.992	.986	.937	.990	.963	95.40
		M^{3}SDA [10]	.951	.919	.935	.964	.939	.951	.969	.943	.956	.973	.994	.984	.938	**.992**	.964	95.86
	DG	Disentangle [7]	**.956**	.935	.945	.957	.927	.942	.933	.953	.943	.987	.982	.985	.960	.991	.975	95.85
	MCDCD	MCDCD (w/o $\mathcal{L}_{trip}$)	.938	.971	.954	.956	.967	.962	**.976**	**.955**	**.965**	**.992**	.989	.991	.987	.976	.982	97.10
		MCDCD	.942	**.988**	**.964**	**.965**	**.972**	**.968**	.973	.954	.964	.985	**.999**	**.993**	**.995**	.975	**.985**	**97.61**
EMG	Baselines	BiLSTM	.152	.151	.152	.379	.335	.356	.387	.441	.412	.497	.474	.485	.441	.454	.447	37.86
		BiLSTM_attn	.184	.150	.165	.369	.317	.341	.374	.424	.398	.514	.518	.516	.472	.539	.503	39.84
		1D-CNN	.181	.143	.159	.378	.289	.328	.427	**.482**	**.453**	.534	.543	.538	.494	.614	.547	42.64
	SS-UDA	DANN [38]†	.352	.195	.251	.376	.335	.354	.437	.439	.438	.452	.581	.508	.497	.595	.541	42.19
		DANN [38]$^{\#}$	.274	.206	.235	.367	.322	.343	.378	.450	.411	.510	.509	.509	.479	.539	.507	41.41
		MCD [11]†	.377	.287	.325	.396	.345	.368	.402	.315	.353	.495	.622	.551	.500	.654	.567	43.78
		MCD [11]$^{\#}$	.304	.292	.298	.395	.353	.373	.426	.425	.425	.504	.524	.514	.494	.543	.517	43.26
	MS-UDA	MDAN [35]	.331	**.296**	**.313**	.383	.359	.371	.421	.422	.422	.517	.529	.523	.485	.542	.512	43.59
		MS-MCD	**.484**	.205	.288	.371	.414	.391	.389	.305	.342	.563	**.740**	.639	.553	.690	.613	48.41
		M^{3}SDA [10]	.330	.139	.196	**.411**	.513	**.456**	.413	.298	.346	.553	.730	.630	.575	.667	.618	48.40
	DG	Disentangle [7]	.227	.221	.224	.358	**.601**	.445	**.511**	.333	.403	**.607**	.486	.540	.551	.493	.520	43.44
	MCDCD	MCDCD (w/o $\mathcal{L}_{trip}$)	.446	.180	.256	.412	.456	.432	.431	.397	.413	.594	.724	**.653**	.570	**.703**	.629	50.32
		MCDCD	.458	.185	.263	.398	.443	.419	.437	.439	.438	.569	.730	.639	**.636**	**.703**	**.668**	**51.32**

• (U)DA - (Unsupervised) Domain Adaptation; SS - Single Source; MS - Multi-Source; DG - Domain Generalization; • Elements in **bold** and underline indicate the best and second best in each column, respectively. • For SS-UDA methods (DANN and MCD), † indicates the results using *single best* standard, and $^{\#}$ is the one with *source combine* standard.



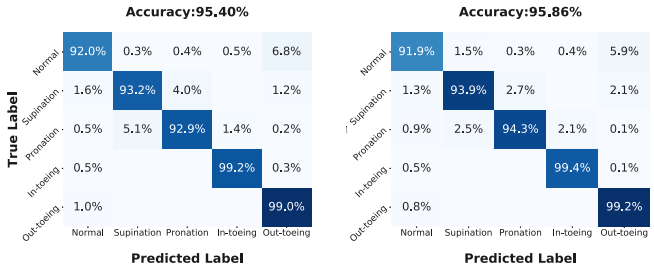
(a) Mocap (b) EMG

Fig. 6. Evaluation of the hyperparameter $n_{iter} = \{1, 2, 3, 4, 5, 6\}$ for the proposed MCDCD and the state-of-the-art UDA methods.

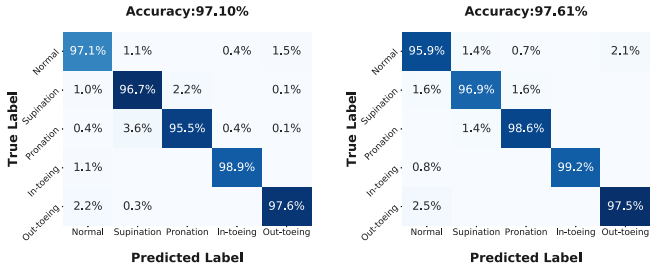
observation emphasizes that the proposed cross-domain classifier discrepancy loss can not only improve the performance in abnormal gait recognition but also facilitate faster convergence speed (smaller n_{iter}) of the domain adaptation across multiple sources.

2) Results on Mocap Data: As shown in Table II, in terms of the high-quality Mocap data, all the methods achieved satisfactory results in subtle gait abnormality detection. For single-source UDA methods (i.e., DANN & MCD), it can be seen that results with *source combine*(#) standard outperform those with *single best*(†). Moreover, due to the existence of individual differences, it can be found that the performance by the single-source UDA methods is similar to baseline methods, which reveals the limitation of taking multiple subjects as one source domain. On the other hand, we can also observe that all multi-source UDA methods (MDAN, MS-MCD, & M ^{3} SDA) outperform the baselines and single-source counterparts, where the improvement of Acc is around 3%. Our proposed MCDCD achieves the best result on abnormal gait recognition (Acc = 97.61%), which outperform the second best methods (MS-MCD, M ^{3} SDA [10], and Disentangle [7]) by 1.7%∽2.2%. Besides, prec, rec, and F1 by MCDCD for each gait category all reach either the best or the second best among all the methods. Actually, the major improvement come from the better classification of {Normal, Supination, Pronation}, where these three patterns are difficult to discriminate. As can be seen from the confusion matrices in Fig. 7, the accuracy of these three classes increases from around 93% (MS-MCD & M ^{3} SDA) to 97%. Without the use of triplet loss, the overall recognition accuracy is decreased, which mainly originates from the degraded performance of Supination class (see Fig. 7(c)). Meanwhile, taking a comparison of MS-MCD and MCDCD (w/o $\mathcal{L}_{trip}$), it also illustrates the effectiveness of the proposed cross-domain discrepancy strategy.

Fig. 9(a) visualizes the Mocap feature distribution of the proposed MCDCD during the training procedure through the use of t-distributed Stochastic Neighbor Embedding (t-SNE). For better visualization, we only demonstrate the features from four source domains (training subjects). Before the training (see top row), the feature distributions across subjects are entangled with each other, especially for the gait patterns {Normal, Supination, Pronation}. As shown in the bottom row, after the training with the proposed MCDCD, the domain shift between any two source domains can be eliminated. More importantly, with the

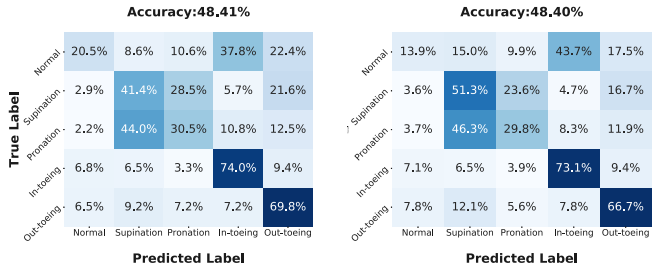


(a) MS-MCD (b) M^3SDA

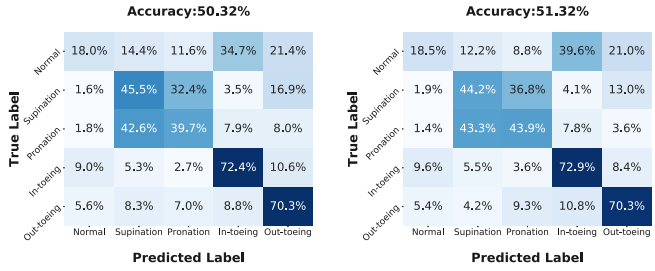


(c) MCDCD (w/o $\mathcal{L}_{trip}$) (d) MCDCD

Fig. 7. Confusion matrices of different MS-UDA methods on **Mocap** data.



(a) MS-MCD (b) M^3SDA

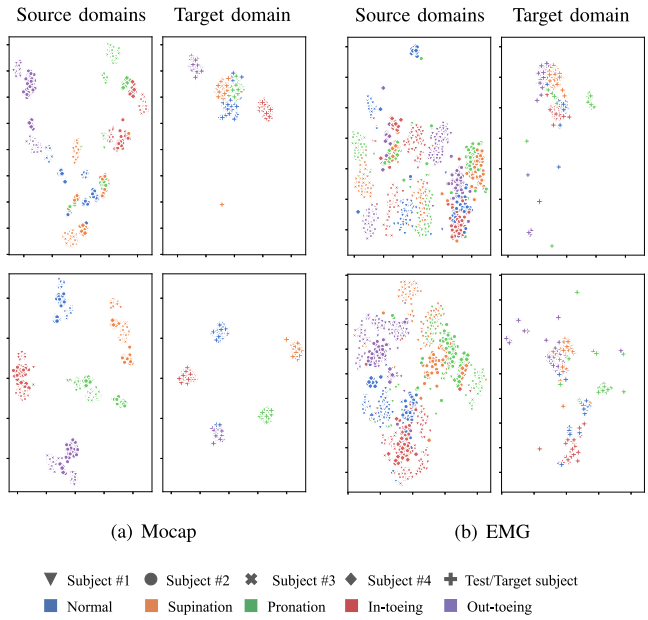


(c) MCDCD (w/o $\mathcal{L}_{trip}$) (d) MCDCD

Fig. 8. Confusion matrices of different MS-UDA methods on **EMG** data.

help of triplet loss, the intra-class features are clustered and the inter-class features are separated for both source and target domains.

3) Results on EMG Data: As EMG data is always contaminated by various noise factors, the recognition accuracies using EMG data are significantly degraded compared to those using Mocap data. For single-source UDA methods, the performance on both *single best* and *source combine* standards achieves similar recognition results. However, a similar conclusion can be drawn that the multi-source UDA methods outperform the baseline networks and those with single-source UDA. The best result is still achieved by our MCDCD ($\text{Acc} = 51.32\%$). As demonstrated by the results in Table II and confusion matrices as in Fig. 8, the Normal gait pattern is the most difficult to discriminate. Meanwhile, similar to the Mocap data, the Supination and Pronation patterns are entangled with the Normal gait pattern, leading to unsatisfactory classification rates. From the t-SNE plot of the EMG data in Fig. 9(b), there exist large discrepancies among the feature distribution of different source domains, which significantly increase the difficulty for diminishing these shifts as well as clustering intra-class features. After training with MCDCD (see bottom row of Fig. 9(b)), the distributions of source domains and the target domain are aligned, however, the inter-class samples are still entangled with each other, while. As a result, the abnormal gait recognition solely using EMG data is still challenging.



(a) Mocap (b) EMG

Fig. 9. t-SNE visualization of the (a)-**Mocap** or (b)-**EMG** feature distribution. Top row is the features before unsupervised domain adaptation; Bottom row shows the feature distribution after training by MCDCD. Features of the source and target domains are displayed within the same range.

D. ANOVA of Recognition Results

In order to highlight the effectiveness of the proposed MCDCD, we performed the statistical analysis of the recognition results between each compared state-of-the-art method and our proposed MCDCD. As shown in Fig. 10, we use the marker * and ** to indicate $0.005 \leq p < 0.05$ and $p < 0.005$, where p-values derived from one-way ANOVA. It can be observed that In terms of both Mocap and EMG data, it can be observed that there exist statistically significant differences for both Mocap and EMG data.

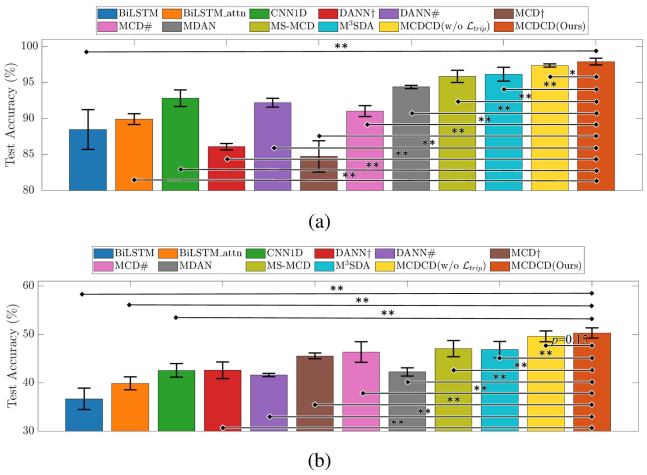


Fig. 10. ANOVA results between each method and the proposed MCDCD. (a) Mocap data; (b) EMG data. The marker * and ** indicating $0.005 \leq p < 0.05$ and $p < 0.005$, respectively.

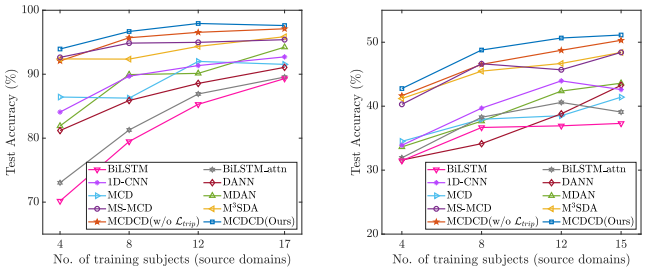


Fig. 11. Comparison results on the case using a small number of training subject (source domains) to train the model.

E. Results on Small Number of Training Subjects

One of the inherent challenges in healthcare applications is the lack of large-scale ground truth data. Hence, it is critical that the computational model can generalize to new subjects when trained on a small-size dataset. For our abnormal gait dataset, there exist 18 subjects and we used data from 17 subjects for training in each run. To validate the robustness of the proposed method, we additionally conducted the comparison experiments on a more challenging case with a smaller number of the training subjects (source domains) as shown in Fig. 11. As can be seen from the top row (Mocap data), with a small number of training subjects, there is clear evidence that the multi-source UDA methods outperform the single-source ones, where the baseline networks are inferior to the UDA methods. For EMG data as presented in the bottom row, the recognition accuracies of various methods slightly drop with a small number of source domains. Nevertheless, the proposed MCDCD still achieves the best performance, which reveals its effectiveness and robustness in more challenging cases.

F. Fine-Tuned Results on EMG Data

Although the proposed MCDCD method achieves superior performance in abnormal gait recognition compared to other

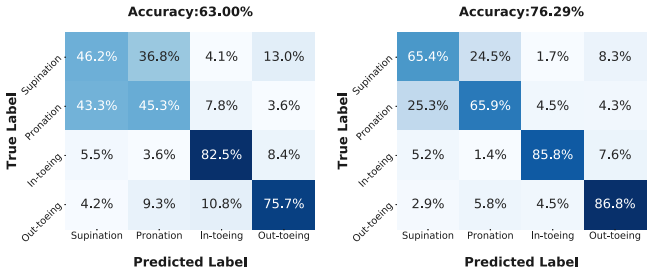


Fig. 12. Comparison results on four abnormal classes recognition using EMG data. (a) Original best result by MCDCD (corresponding to 51.32% for 5 classes in Table II & Fig. 8(d)); (b) Improved result on EMG data by exploiting the Normal gait pattern of the target/test subject as an additional source domain.

UDA approaches, the results using EMG data are still not satisfactory. This is an inherent challenge for EMG data, as even for the same anatomical localization, the recorded EMG waveform would shape totally differently across subjects. To demonstrate the potential improvement of the proposed MCDCD for EMG-based abnormal gait recognition, we conduct the experiment by taking out the **Normal** data of the target domain $\mathcal{D}_T$ as a new source domain $\mathcal{D}_{N+1}$, where the remaining four abnormal patterns {In-toeing, Out-toeing, Supination, Pronation} are still regarded as the target domain. In this manner, we compared this fine-tuned result with the original best (MCDCD - 51.32% for 5 abnormal classes; 63.00% for 4 abnormal classes as in Fig. 12(a)). It can be observed from Fig. 12(b) that the recognition accuracy increases to 76.29%, which exceeds the original best by around 13%. This observation implies that the pre-collection of the **Normal** EMG data could be beneficial for detecting the abnormal gait of the same subject afterwards.

VI. Conclusion

To address the inherent challenges of individual differences in gait analysis, we proposed a novel framework, namely MCDCD, in the context of multi-source unsupervised domain adaptation, which improves the abnormal gait recognition performance on a novel subject (target domain). The proposed network is composed of a single feature extractor and domain-specific category classifiers. The core idea is to diminish the domain shift between source domains (training subjects) and the target subject by leveraging the information from multiple sources. We conducted extensive comparison experiments on with our abnormal gait dataset. Experiment results have demonstrated the effectiveness of MCDCD in recognizing gait abnormality on a novel subject, especially illustrating the contribution of the unsupervised domain adaptation with the cross-domain classifier discrepancy and the triplet loss.

Future work will target on the cross-modal transfer learning to extract the discriminative features from the noisy modality by leveraging the information from high-quality Mocap data. Beyond the multi-source domain adaptation, there is also a pressing need for proposing novel frameworks to address the multi-target unsupervised domain adaptation, which aims to

enhance the generalization capability of deep models in terms of multiple novel subjects.

REFERENCES

[1] P. Mahlknecht *et al.*, "Prevalence and burden of gait disorders in elderly men and women aged 60-97 years: A population-based study," *PLoS One*, vol. 8, no. 7, 2013, Art. no. e69627.

[2] A. Lundberg, O. Svensson, C. Bylund, I. Goldie, and G. Selvik, "Kinematics of the ankle/foot complex-part 2: Pronation and supination," *Foot Ankle Int.*, vol. 9, no. 5, pp. 248–253, 1989.

[3] P. Sass and G. Hassan, "Lower extremity abnormalities in children," *Amer. Fam. Physician*, vol. 68, no. 3, pp. 461–468, 2003.

[4] F. Deligianni, Y. Guo, and G.-Z. Yang, "From emotions to mood disorders: A survey on gait analysis methodology," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2302–2316, Nov. 2019.

[5] S. Chen, J. Lach, B. Lo, and G.-Z. Yang, "Toward pervasive gait analysis with wearable sensors: A systematic review," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 6, pp. 1521–1537, Nov. 2016.

[6] D. Raví *et al.*, "Deep learning for health informatics," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 4–21, 2016.

[7] X. Gu, Y. Guo, F. Deligianni, B. Lo, and G. -Z. Yang, "Cross-subject and cross-modal transfer for generalized abnormal gait pattern recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 546–560, Jan. 2020.

[8] W. M. Kouw and M. Loog, "A review of domain adaptation without target labels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 766–785, Mar. 2021.

[9] J. Hoffman, M. Mohri, and N. Zhang, "Algorithms and theory for multiple-source adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 8246–8256.

[10] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 1406–1415.

[11] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3723–3732.

[12] G.-Z. Yang, *Body Sensor Networks*. vol. 1. Berlin, Germany: Springer, 2014.

[13] S. Bei, Z. Zhen, Z. Xing, L. Taocheng, and L. Qin, "Movement disorder detection via adaptively fused gait analysis based on kinect sensors," *IEEE Sensor J.*, vol. 18, no. 17, pp. 7305–7314, Sep. 2018.

[14] Y. Guo, F. Deligianni, X. Gu, and G.-Z. Yang, "3-D canonical pose estimation and abnormal gait recognition with a single RGB-D camera," *IEEE Robot. Automat. Lett.*, vol. 4, no. 4, pp. 3617–3624, Oct. 2019.

[15] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," *Sports Med.-Open*, vol. 4, no. 1, pp. 1–15, 2018.

[16] T. Seel, J. Raisch, and T. Schauer, "IMU-based joint angle measurement for gait analysis," *Sensors*, vol. 14, no. 4, pp. 6891–6909, 2014.

[17] D. Jarchi, C. Wong, R. M. Kwasnicki, B. Heller, G. A. Tew, and G. Yang, "Gait parameter estimation from a miniaturized ear-worn sensor using singular spectrum analysis and longest common subsequence," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1261–1273, Apr. 2014.

[18] D. Jarchi, B. Lo, C. Wong, E. Ieong, D. Nathwani, and G.-Z. Yang, "Gait analysis from a single ear-worn sensor: Reliability and clinical evaluation for orthopaedic patients," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 8, pp. 882–892, Aug. 2016.

[19] F. Mu, X. Gu, Y. Guo, and B. Lo, "Unsupervised domain adaptation for position-independent imu based gait analysis," in *Proc. IEEE Sensors*, 2020, pp. 1–4.

[20] V. Agostini, M. Ghislieri, S. Rosati, G. Balestra, and M. Knaflitz, "Surface electromyography applied to gait analysis: How to improve its impact in clinics?," *Front. Neurol.*, vol. 11, no. 994, pp. 1–13, 2020.

[21] A. Phinyomark, A. Nuidod, P. Phukpattaranont, and C. Limsakul, "Feature extraction and reduction of wavelet transform coefficients for emg pattern classification," *Electron. Elect. Eng.*, vol. 122, no. 6, pp. 27–32, 2012.

[22] S. W. Lee, T. Yi, J. Jung, and Z. Bien, "Design of a gait phase recognition system that can cope with EMG electrode location variation," *IEEE Trans. Automat. Sci. Eng.*, vol. 14, no. 3, pp. 1429–1439, Jul. 2017.

[23] R. Luo, S. Sun, X. Zhang, Z. Tang, and W. Wang, "A low-cost end-to-end semg-based gait sub-phase recognition system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 267–276, Jan. 2020.

[24] Y. Guo, R. Gravina, X. Gu, G. Fortino, and G.-Z. Yang, "EMG-based abnormal gait detection and recognition," in *Proc. IEEE Int. Conf. Hum.-Mach. Syst.*, 2020, pp. 1–6.

[25] J. He, K. Lippmann, N. Shakoor, C. Ferrigno, and M. A. Wimmer, "Unsupervised gait retraining using a wireless pressure-detecting shoe insole," *Gait Posture*, vol. 70, pp. 408–413, 2019.

[26] T.-N. Nguyen, H.-H. Huynh, and J. Meunier, "Skeleton-based abnormal gait detection," *Sensors*, vol. 16, no. 11, 2016, Art. no. 1792.

[27] Y. Feng, Y. Li, and J. Luo, "Learning effective gait features using lstm," in *Proc. 23rd Int. Conf. Pattern Recognit.*, 2016, pp. 325–330.

[28] L. Kidzinski, S. L. Delp, and M. H. Schwartz, "Automatic real-time gait event detection in children using deep neural networks," *PLoS One*, vol. 14, no. 1, 2019, Art. no. e0211466.

[29] J. Hannink, T. Kautz, C. Pasluosta, K. G. Gasmann, J. Klucken, and B. M. Eskofier, "Sensor-based gait parameter extraction with deep convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 85–93, Jan. 2017.

[30] K. Hu *et al.*, "Graph sequence recurrent neural network for vision-based freezing of gait detection," *IEEE Trans. Image Process.*, vol. 29, pp. 1890–1901, Oct. 2020.

[31] J. Gao, P. Gu, Q. Ren, J. Zhang, and X. Song, "Abnormal gait recognition algorithm based on LSTM-CNN fusion network," *IEEE Access*, vol. 7, pp. 163180–163190, 2019.

[32] F. Horst, S. Lapuschkin, W. Samek, K. Muller, and W. I. Schollhorn, "Explaining the unique nature of individual gait patterns with deep learning," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 2391.

[33] S. Bendavid, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, no. 1, pp. 151–175, 2010.

[34] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by back-propagation," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[35] H. Zhao, S. Zhang, G. Wu, J. M. Moura, J. P. Costeira, and G. J. Gordon, "Adversarial multiple source domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 8559–8570.

[36] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv:1412.3474*, 2014. [Online]. Available: https://arxiv.org/abs/1412.3474.

[37] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2208–2217.

[38] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 189–209, 2016.

[39] J. Hoffman *et al.*, "CyCADA: Cycle-consistent adversarial domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1989–1998.

[40] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1857–1865.

[41] R. Xu, Z. Chen, W. Zuo, J. Yan, and L. Lin, "Deep cocktail network: Multi-source unsupervised domain adaptation with category shift," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3964–3973.

[42] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," *arxiv:1703.07737*, 2017. [Online]. Available: https://arxiv.org/abs/1703.07737.

[43] Y. Guo, Y. Li, and Z. Shao, "DSRF: A flexible trajectory descriptor for articulated human action recognition," *Pattern Recognit.*, vol. 76, pp. 137–148, 2018.

[44] W. Geng *et al.*, "Gesture recognition by instantaneous surface emg images," *Sci. Rep.*, vol. 6, no. 36571, pp. 1–8, 2016.