Unsupervised Cross-Subject Adaptation for Predicting Human Locomotion Intent

Kuangen Zhang, Jing Wang, Clarence W. de Silva, *Life Fellow, IEEE*, and Chenglong Fu

***Abstract*— Accurately predicting human locomotion intent is beneficial in controlling wearable robots and in assisting humans to walk smoothly on different terrains. Traditional methods for predicting human locomotion intent require collecting and labeling the human signals, and training specific classifiers for each new subject, which introduce a heavy burden on both the subject and the researcher. In addressing this issue, the present study liberates the subject and the researcher from labeling a large amount of data, by incorporating an unsupervised cross-subject adaptation method to predict the locomotion intent of a target subject whose signals are not labeled. The adaptation is realized by designing two classifiers to maximize the classification discrepancy and a feature generator to align the hidden features of the source and the target subjects to minimize the classification discrepancy. A neural network is trained by the labeled training set of source subjects and the unlabeled training set of target subjects. Then it is validated and tested on the validation set and the test set of target subjects. Experimental results in the leave-one-subject-out test indicate that the present method can classify the locomotion intent and activities of target subjects at the averaged accuracy of 93.60% and 94.59% on two public datasets. The present method increases the user-independence of the classifiers, but it has been evaluated only on the data of subjects without disabilities. The potential of the present method to predict the locomotion intent of subjects with disabilities and control the wearable robots will be evaluated in future work.**

Manuscript received October 1, 2019; revised November 19, 2019; accepted January 7, 2020. Date of publication January 15, 2020; date of current version March 6, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1305400, in part by the National Natural Science Foundation of China under Grant U1913205, Grant 61533004, and Grant U1613206, in part by the Guangdong Innovative and Entrepreneurial Research Team Program under Grant 2016ZT06G587, in part by the Shenzhen and Hong Kong Innovation Circle Project under Grant SGLH20180619172011638, and in part by the Centers for Mechanical Engineering Research and Education at MIT and SUSTech. *(Kuangen Zhang and Jing Wang contributed equally to this work.) (Corresponding author: Chenglong Fu.)*

Kuangen Zhang is with the Department of Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen 518055, China, and also with the Department of Mechanical Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

Jing Wang and Clarence W. de Silva are with the Department of Mechanical Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

Chenglong Fu is with the Department of Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: fucl@sustech.edu.cn).

Code and data: https://github.com/KuangenZhang/UCSA_HIP

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2020.2966749

***Index Terms*— Cross-subject adaptation, unsupervised learning, human intent classification, wearable robots.**

I. Introduction

INTENT is a mental activity that indicates a commitment to carry out specific actions in the future. Predicting the intent of other individuals is important for both animals and robots. Animals can predictively adjust their actions after predicting the internal mental states and potential future actions of prey, predators, and mates [1]. Robots would coordinate with other individuals better if they could accurately predict the intent of other entities, especially humans. For instance, powered wearable robots (e.g., prostheses, orthoses, extra robots, and exoskeletons) [2]–[5] need to predict human locomotion intent to predictively change their locomotion modes to adapt to different terrains (eg., stairs, ramp, and level ground) [6], [7]. Researchers [8]–[11] claim that an intent recognition system is critical for the wearable robots in order to avoid disrupting the gait cycle and adapt to different locomotion modes. If wearable robots cannot predict the human locomotion intent accurately and timely, wearable robots may switch to the level-ground mode while climbing stairs, which would disrupt the gait cycle of humans and might cause the humans to tumble.

It is difficult for a robot to understand human intent, which cannot be measured directly. To address this issue, previous researchers captured the signals in the human-robot interface, including the signals of the inertial measurement unit (IMU) [10]–[13], the surface electromyography (EMG) signals of the residual limb [14], and the pressure as recorded from the pressure-sensitive insoles [15]. Some researchers directly controlled joint angles or torques of wearable robots based on the captured human signals [2]. However, such volitional control methods are not robust. Most researchers prefer to classify these signals to some motion patterns, for switching the locomotion modes of the wearable robots via a finite-state controller [7]. Typical classification methods include linear discriminant analysis (LDA), support vector machines (SVM) [16], [17], and artificial neural networks (ANN) [17], [18]. LDA is computationally efficient because it can be solved analytically, but it cannot handle complex features. SVM can deal with complex features but a suitable kernel function should be designed manually. ANN requires less experience, but its classification accuracy based on ANN may be lower than LDA and SVM [9], [19]. Compared to shallow classification methods, deep neural networks, such as a convolutional neural network (CNN) [20], can classify human intent more accurately and do not rely on the human experience [21]. However, the computational complexity of deep neural networks is greater than that of shallow classification methods. A common limitation for all these methods is that they all require labeled signals for every subject because signals in the human-robot interface are usually noisy and user-dependent. For instance, these signals can be affected by the sensor position and the skin quality of humans. To accurately decode the human locomotion intent, researchers need to collect and label the signals and train the classifier for each new subject, which is burdensome for both the subject and the researcher [9].

Considering that the signals in the human-robot interface are usually user-dependent, some researchers utilized a vision sensor to monitor the environment and estimate the locomotion modes [22], [23]. Besides, the trajectory of the robot can be predicted from the sequential images by a novel self-supervised learning method, which is able to supply latent representations with physical semantic meanings for controlling the robot [24]. In our previous study, environmental information captured by a depth camera and an IMU was utilized to classify the environment and estimate the environmental parameters. The environmental classifier trained for one subject achieved similar classification accuracy for other subjects without training, which was user-independent [25]. Nevertheless, the environmental classification cannot determine the locomotion intent accurately, such as the accurate transition time between two locomotion modes. To accurately predict the human locomotion intent, signals in the human-robot interface are still required.

To design a user-independent classifier to accurately classify the user-dependent signals, some researchers have adopted domain adaptation strategies. They assumed that signals were from two different types of subjects: the source subjects and the target subjects. Sufficient labeled signals are available from the source subjects, while it is difficult to label the signals of the target subjects. To address this issue, researchers designed unsupervised domain adaptation methods to train the classifier using the labeled data from the source subjects and the unlabeled data from the target subjects. They then used the trained classifier to classify the data of the target subjects [26]. Although the feasibility of unsupervised domain adaptation methods has been validated, there are still some limitations. First, the focus had been on recognizing human activities using the IMU signals with delay rather than predicting human locomotion intent. Human activities can be observed directly while human locomotion intent happens mentally and cannot be observed. Hence, it is more difficult to accurately predict human locomotion intent. In addition, the state-of-the-art accuracy (87%) for recognizing the activities of the target subject is not satisfactory for practical application. Moreover, the domain adaptation method proposed in [26] was based on an unsupervised clustering method, which required multiple iterations and could take a longer time than an end-to-end domain adaptation method.

Many end-to-end domain adaptation methods have been proposed to classify images, including domain-adversarial neural networks (DANN) [27], domain separation networks (DSN) [28], and adversarial discriminative domain adaptation (ADDA) method [29]. The most representative method is DANN, which consists of a feature generator, a label classifier, and a domain classifier. The feature generator generates the hidden features, which are input to the domain classifier to classify their domains and to the label classifier to classify their labels. The feature generator aligns features from the source domain and the target domain to fool the domain classifier until it can not discern which domain the features came from. However, there are some limitations. First, the domain classifier does not consider the classes of the target domain samples, and thus a trained generator may generate ambiguous features near class boundaries. Second, the feature distributions cannot be aligned entirely between different domains due to different characteristics of different domains. Besides, the above methods have only been evaluated on processing images, which are usually more stable than human signals.

In response to such limitations and inspired by an image classification method proposed by Saito *et al.* [30], this paper proposes an end-to-end unsupervised cross-subject adaptation method to accurately classify the locomotion intent and activities of the target subject based on the human kinetic and biological signals (e.g., IMU and EMG) (see Fig. 1). To the best of our knowledge, predicting human locomotion intent without labeling the signals from the target subjects has not been resolved to date. To realize the cross-subject adaptation, a convolutional network, which consists of one feature generator and two classifiers, is designed to align the hidden features and classify the aligned features in an adversarial manner. Because one important step of training the network is to maximize the classification discrepancy between two classifiers, which is similar to the training strategy proposed by Saito *et al.* [30], this method is also named as MCD. It is hypothesized that the designed network trained by the labeled data from the source subjects and the unlabeled data from the target subjects can still predict the locomotion intent of the target subjects accurately.

The key contributions of the present paper include the following:

1) Development of an end-to-end unsupervised cross-subject adaptation method to predict the locomotion intent of the target subjects accurately and efficiently.
2) Designing a novel CNN to align the features of the source subjects and the target subjects in an adversarial manner.
3) Evaluating the developed method on two public datasets and achieving state-of-art accuracy (93.60% and 94.59%) for classifying the locomotion intent and activities of the target subjects.
4) Comparing the performance of different sensors and identifying the most important sensor to classify human locomotion intent and activities.

The rest of this paper is organized as follows. Section II describes the theoretical methods, the network architecture, and the experimental setup. Sections III and IV present the

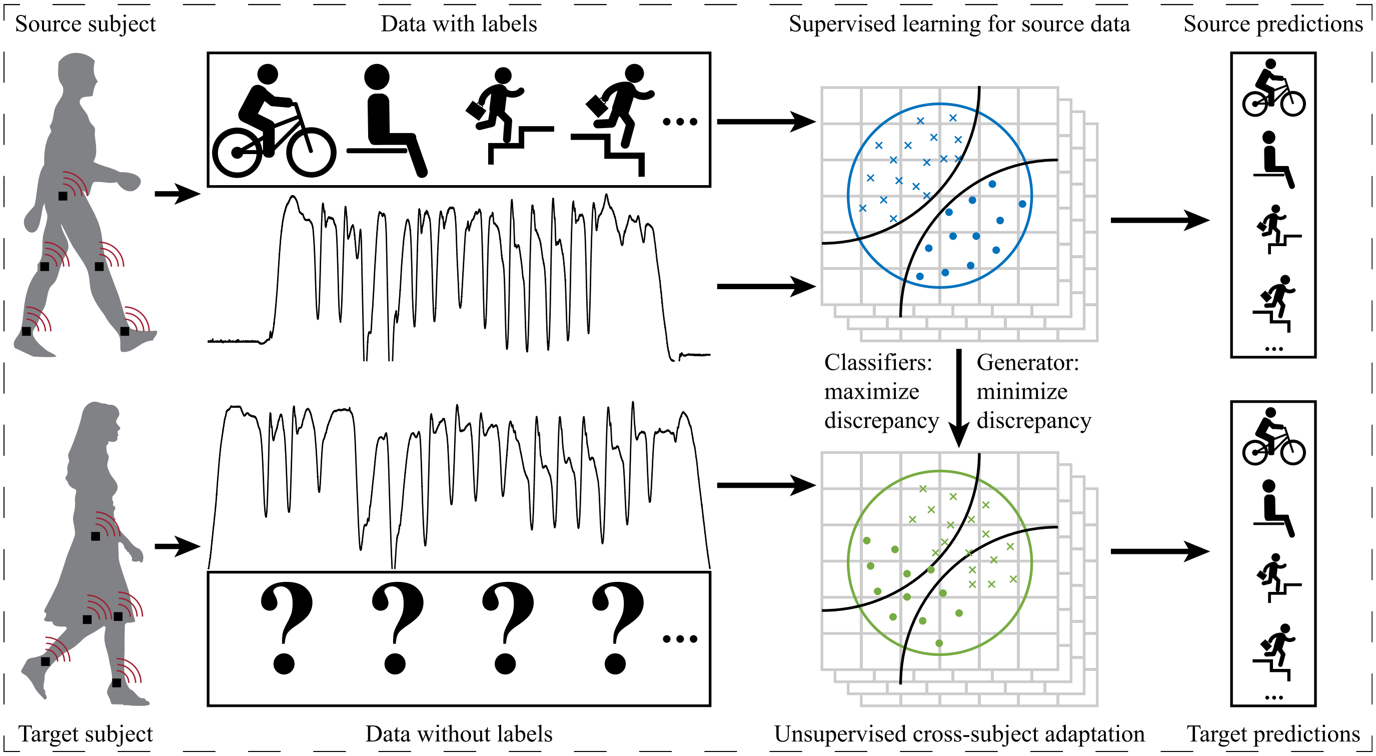

Source subject
Data with labels
Supervised learning for source data
Source predictions
Classifiers: maximize discrepancy
Generator: minimize discrepancy
? ? ? ?...
Target subject
Data without labels
Unsupervised cross-subject adaptation
Target predictions

Fig. 1. **The overview of the unsupervised cross-subject adaptation method.** The data of source subjects are labeled and can be trained by the supervised learning method. In contrast, the data of target subjects do not have labels and are trained by the unsupervised cross-subject adaptation method, which is realized by designing two classifiers to maximize the discrepancy and a feature generator to minimize the discrepancy between two classifiers. Finally, the proposed method can predict labels of both source data and target data.

experimental results and discuss them. Section V concludes the paper.

II. Materials and Methods

This section describes the proposed methods in detail, including the signal processing method and unsupervised domain adaptation method. The proposed methods are evaluated using two public datasets, and the corresponding experimental setup and statistical analysis methods are presented.

A. Signal Processing

The present paper processes two public datasets: the encyclopedia of able-bodied bilateral lower limb locomotor signals (ENABL3S)[1] provided by the Northwestern University [31] and the daily and sports activities data set (DSADS)[2] provided by the Bilkent University [32]. The signals in these datasets have been filtered and segmented, and detailed signal processing methods have been introduced in their papers. To keep the integrity of the present paper, the signal segmentation and feature extraction are introduced briefly in this section.

The ENABL3S [31] includes filtered signals of bilateral EMG, IMU, and joint angle sensors. These filtered signals were segmented by 300 ms wide sliding windows. The sliding windows began 300 ms before each gait event, such as heel contact and toe-off. It is difficult to accurately define and measure the accurate time of human intent, which is not an intuitive signal. Hence, previous researchers had to utilize some gait events to estimate the time of switching between different locomotion modes of the wearable robots, such as the toe-off from the level-ground mode to the upstairs mode and the heel contact from downstairs mode to the level-ground mode [11], [25], [33]. In the present work, locomotion modes can be classified before the gait event, which is used to trigger the transition of the locomotion modes for the wearable robots. The features were extracted from the segmented signals. For EMG signals, ten features, including waveform length, the mean absolute value, the coefficients of a sixth-order autoregressive model, the number of slope sign changes, and the number of zero crossings, were extracted. Features for the IMU and the joint angular signals included the maximum, minimum, mean, standard deviation, and initial and final values. In the present study, the features of 14 EMG sensors, 5 IMUs, and 4 joint angle sensors were reshaped to a 33×12 matrix, and each row represents features of a sensor.

The DSADS [32] contains five 9-axis IMUs and the signals were captured at 25 Hz. The captured signals were segmented to 5 s segments. There were 45 signal channels, and extracted features from each channel included the maximum, minimum, mean, standard deviation, and initial and final values. Therefore, a 45×6 feature matrix was formed. There was no transition between different activities and the segmented signals were delayed, and thus this dataset was only used to compare the performance of recognizing human activities using the

[1] https://doi.org/10.6084/m9.figshare.5362627

[2] http://archive.ics.uci.edu/ml/datasets/Daily+and+Sports+Activities

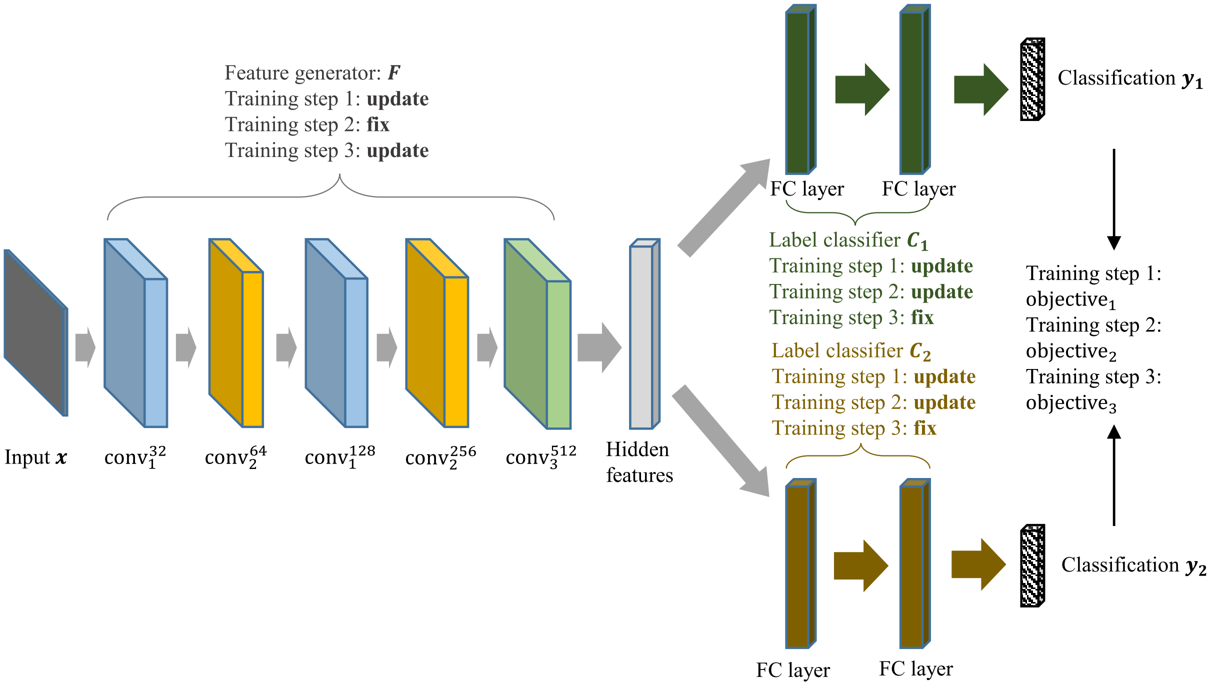

Feature generator: $\boldsymbol{F}$
Training step 1: **update**
Training step 2: **fix**
Training step 3: **update**
Input $\boldsymbol{x}$
conv_1^{32}
conv_2^{64}
conv_1^{128}
conv_2^{256}
conv_3^{512}
Hidden features
FC layer
FC layer
Classification $\boldsymbol{y_1}$
Label classifier $\boldsymbol{C_1}$
Training step 1: **update**
Training step 2: **update**
Training step 3: **fix**
Label classifier $\boldsymbol{C_2}$
Training step 1: **update**
Training step 2: **update**
Training step 3: **fix**
Training step 1: objective$_1$
Training step 2: objective$_2$
Training step 3: objective$_3$
FC layer
FC layer
Classification $\boldsymbol{y_2}$

Fig. 2. The proposed network architecture and training steps. The type of convolutional layer used is indicated by both of the subscript and the superscript, e.g., conv_1^{32} suggests that a type-one convolutional layer with 32 channels is used. The fully connected layer is denoted by FC. Training steps and objectives are introduced in subsubsection II-B.3 and 3-5.

present method and that using the methods introduced in the previous study [26].

B. Unsupervised Cross-Subject Adaptation Network

The developed network is based on the network of maximum classifier discrepancy (MCD) [30] and a theorem proposed by Ben-David *et al.* [34] on bounding the conditions where a classifier trained from the source data can be expected to perform well on the target data. Their work suggests that the target error ϵ_T of a classifier can be upper bounded by the sum of the source error ϵ_S of a classifier, the distance in the symmetric difference classifier space between the two classifiers $d_{H\Delta H}(S,T)$, and the combined error of the ideal joint classifier λ:

$$\epsilon_T(h) \leq \epsilon_S(h) + \frac{1}{2} d_{H\Delta H}(D_S, D_T) + \lambda, \tag{1}$$

where h denotes a classifier that belongs to a classifier space H. D_S and D_T represent the source domain and the target domain, respectively.

The $d_{H\Delta H}(S,T)$ is defined as [34]:

$$\begin{aligned} d_{H\Delta H}(D_S, D_T) = & 2 \sup_{h,h' \in H} |Pr_{x\sim D_S}[h(x) \neq h'(x)] \\ & - Pr_{x\sim D_T}[h(x) \neq h'(x)]|, \end{aligned} \tag{2}$$

where sup indicates the upper bound, and Pr denotes the probability. Two classifiers h and h' belong to the classifier space H.

The source error ϵ_S can be minimized by training h and h' using the same labeled data in the source domain. The term $Pr_{x\sim D_S}[h(x) \neq h'(x)]$ inside $d_{H\Delta H}(S,T)$ will also be theoretically minimum, and can be assumed to be unimportant under the setting of unsupervised domain adaptation. As a result, the objective of this research is to design a learning algorithm to train two different classifiers h and h' to agree on the predictions in the target domain, and make $Pr_{x\sim D_T}[h(x) \neq h'(x)]$ close to $Pr_{x\sim D_S}[h(x) \neq h'(x)]$.

Inspired by the definition of $d_{H\Delta H}(D_S, D_T)$, the present paper designs two classifiers C_1 and C_2 with the same fully connected neural network architecture and builds a convolutional neural network F to be the feature generator whose outputs are shared by two different classifiers. Thus, two classifiers h and h' mentioned above are defined as $h = C_1[F(x)]$ and $h' = C_2[F(x)]$ in this research. Additionally, the objective becomes how to jointly train h and h' to realize $Pr_{x\sim D_T}[h(x) \neq h'(x)] \approx Pr_{x\sim D_S}[h(x) \neq h'(x)]$. Next, this section will discuss the detail of the proposed architectures and the training steps for the feature generator and the label classifiers. The overall network architecture and the adaptation training strategies are shown in Fig. 2.

1) Feature Generator: The human signal matrix to be classified is similar but not identical to images. The relationship between signals from different sensors is needed for the predictive model to classify the human intent correctly. As a result, the proposed feature generator takes advantage of the ability of the CNN to localize pixel dependencies but is slightly different from the traditional architecture of the CNN.

There are three types of convolutional layers in the proposed architecture. The first type of the convolutional layer (conv_1) has multiple 1×1 filters and a stride of 1. This kind of layer is a multi-layer perceptron that increases the total number of features. The second type of convolutional layer (conv_2)

consists of 1×3 filters and a stride of 2. This type of layer convolves features from the same sensor to consider the local connectivity of different features. The third convolutional layer (conv3), which has $m \times n$ filters (m is the number of sensors and n is the number of hidden features for each sensor) and a stride of 1, convolves all features from different sensors to a global hidden feature vector. The relationship among all sensors can be considered rigorously through this convolutional layer.

The batch normalization is adopted after each convolution and before the nonlinear activation function (ReLU). Because the human signals gathered by each sensor are important for the human intent classification, the proposed feature generator does not downsample the features.

2) Label Classifier: The label classifier for the proposed work is an artificial neural network with two fully-connected (FC) layers. One fully-connected layer with batch normalization and a single activation (ReLU) first maps the features extracted from the feature generator to 128 hidden features. Then, another fully-connected layer directly maps these hidden features to the classification scores of N different types of human intent. There are two label classifiers, and their classification scores are summed. The class with the highest score is the most possible class of human intent.

3) Training Steps: The focus of the present work is how to design a learning algorithm to train two different classifiers C_1 and C_2 to agree on their predictions on the target domain. Inspired by the work proposed by Saito *et al.* [30], we solved the problem by separating the learning process into three steps.

Step 1 Based on Ben David *et al.* 's theorem, we first train the feature generator and label classifiers to classify the source signals correctly. This step makes C_1 and C_2 agree on their predictions on the source samples, and thus the source error ϵ_S is minimized through this step. In this training step, the objective 1 is to train the network to minimize the softmax cross entropy loss:

$$\min_{F,C_1,C_2} [\mathbb{E}_{(\boldsymbol{x}_s,y_s)\in(X_s,Y_s)} \sum_{n=1}^{N} -I[n=y_s]\log P_n(y|\boldsymbol{x}_s)], \quad (3)$$

where $I[n = y_s]$ is a binary indicator which is 1 when n equals y_s, P_n denotes the output probability for class n, and $\mathbb{E}$ is the expectation operator.

Step 2 To make the two label classifiers C_1 and C_2 agree on their predictions in the target domain, an adversarial learning strategy is used to detect target features that are far from the support of the source. Thus in this step, the label classifiers are trained as the discriminators on both domains without updating the parameters of the feature generator F. Objective 2 is to maximize the discrepancy between C_1 and C_2 so that the target features without the support of the source can be detected:

$$\min_{C_1,C_2} [\mathbb{E}_{(\boldsymbol{x}_s,y_s)\in(X_s,Y_s)} \sum_{n=1}^{N}(-I[n=y_s]\log P_n(y|\boldsymbol{x}_s)) \\ -\xi\mathbb{E}_{\boldsymbol{x}_t\in X_t}\frac{1}{N}\sum_{n=1}^{N}(|P_n^1(y|\boldsymbol{x}_t)-P_n^2(y|\boldsymbol{x}_t)|)], \quad (4)$$

where P_n^1 and P_n^2 are the output probability from $C_1(F)$ and $C_2(F)$ for class n respectively. ξ is a weight parameter to control the importance of the discrepancy loss.

Step 3 After maximizing the label classifier discrepancy, we then train the feature generator F and fix two label classifiers to only extract target features to make two classifiers achieve an agreement. To better update the feature generator, this training step is repeated four times for the same mini-batch. The objective 3 is described as follows:

$$\min_F[\mathbb{E}_{\boldsymbol{x}_t\in X_t}\frac{1}{N}\sum_{n=1}^{N}|P_n^1(y|\boldsymbol{x}_t)-P_n^2(y|\boldsymbol{x}_t)|]. \quad (5)$$

These three training steps are repeated continuously for different mini-batches until the target error ϵ_T is minimized. After training the network, the classification scores of two classifiers are summed and the class with the highest score is selected as the prediction of the input signals.

C. Implementation Details

The present network was trained by an Adam optimizer, whose learning rate and weight decay were 0.0002 and 0.0005. The max epoch and batch size were 50 and 128, respectively. The dropout rate of all dropout layers was set at 0.5. The network was implemented by PyTorch and tested on a computer with an Intel Core i7-6700K, an 8 GB memory chip (DDR3 SDRAM), and a graphics card (GeForce GTX 1050 Ti).

D. Experimental Setup

For ENABL3S [31], ten able-bodied subjects were invited to perform experiments to capture corresponding human signals. During experiments, these subjects transited between standing (St), level ground walking (LW), stair ascent (SA), stair descent (SD), ramp ascent (RA), and ramp descent (RD). Each subject repeated walking on a circuit ten times, and each circuit consisted of St → LW → SA → LW → RD → LW → St → LW → RA → LW → SD → LW → St. Because transitions between different locomotion modes were recorded, the ENABL3S dataset can be used to predict the locomotion modes.

For DSADS [32], eight able-bodied subjects participated in the experiments and were requested to perform 19 activities (e.g., sitting, standing, running, riding a bike, jumping, and playing basketball). They performed each activity for five minutes and there was no transition between different activities. Hence, the DSADS can only be used to classify human activities rather than predict intent. The DSADS dataset is also selected because it includes more types of activities than in the ENABL3S dataset. There is only one paper that utilizes the domain adaptation strategy to classify human locomotion modes, and it uses the DSADS dataset. Using the same dataset, the present paper can fairly compare the performance of the present method with that of the previous method.

There are about 22,000 signal segments in the ENABL3S dataset and 9,000 signal segments in the DSADS dataset. The data of each subject were randomly divided into the training set (70%), validation set (15%), and the testing (15%) set. In every experiment, a target subject was selected from the subjects and the remaining subjects were regarded as source subjects (leave-one-subject-out test). The network was trained using the labeled training set of source subjects and the unlabeled training set of the target subject. The validation set of the target subject was used to optimize the hyper-parameters of the designed networks and determine the time to stop training. Finally, the performance of the network model was tested on the test set of the target subjects. A different subject was selected as the target subject in the next experiment until transversing all the subjects. The present paper compared the performance of different methods, including LDA, SVM, ANN, CNN, DANN, and the present MCD. The LDA, SVM, and ANN were only trained using the training set of the source subjects and were tested on the test set of the target subject. The network architecture of CNN and MCD are the same but CNN was trained without the training set of the target subject. We did not separately train the classifiers of CNN to maximize the classifier discrepancy nor separately train the feature generator of CNN to minimize the classifier discrepancy. The validation set and the testing set of the target subject were still used to optimize the parameters of the CNN and to test the performance of the CNN. The network architecture of the DANN is also the same as that of the MCD except for the last layer of the domain classifier, which only classifies the domain of the input. The implemented DANN in the present paper uses the same training strategy as that proposed by Ganin *et al.* [27], and is seen as the baseline of the domain adaptation result.

E. Statistical Analysis

In the leave-one-subject-out test, the mean and the standard deviation of the classification accuracy in the source and the target domains were analyzed. The classification accuracy in the source and the target domains was calculated on the test set of the source subjects and the target subject, respectively. Because the classification results followed a normal distribution, a t-test and a one-way ANOVA with post hoc test at a significance level of $P = 0.05$ were used to compare the difference of results between using different methods and between using different sensors. The forward time to classify each signal segment was also calculated to compare the computing time of different methods.

III. Results

A. Evaluation on ENABL3S

On ENABL3S [31], the classification accuracy for each target subject increases significantly with the present MCD. As shown in Fig. 3, the classification accuracy using the MCD is higher than that using LDA, SVM, and ANN for all target subjects, and the MCD also outperforms the CNN and DANN in most cases. Besides, the classification accuracy for different target subjects varies largely because the EMG and IMU signals used in this paper may be user-dependent. The variation of the classification accuracy for different target subjects decreases after using the MCD. The range of the classification accuracy for the target subject decreases to 8.27% after using the MCD, while those using the other methods are higher than 9.22%.

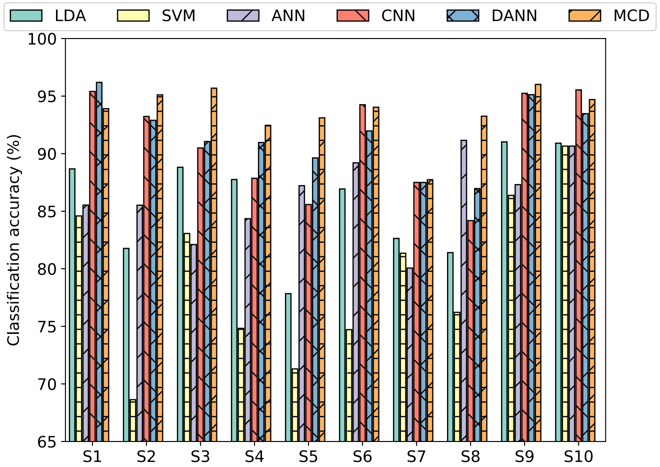


Fig. 3. Accuracy of classifying the locomotion intent for each target subject on ENABL3S using the LDA, SVM, ANN, CNN, DANN, and MCD. S1-S10 represent ten target subjects.

TABLE I

Accuracy and Forward Time of Classifying the Locomotion Intent for the Source Subjects and the Target Subject of ENABL3S Using the LDA, SVM, ANN, CNN, DANN, and MCD Methods. SD Denotes the Standard Deviation. The Forward Time Indicates the Online Executing Time of Classifying Each Signal Segment and the Unit Is in Milliseconds (ms)

Methods	Mean (%)	SD (%)	Mean (%)	SD (%)	Time (ms)
	Source		Target		
LDA	92.59	**0.23**	85.78	4.53	**0.05**
SVM	90.31	1.30	79.17	7.10	**0.05**
ANN	93.56	0.37	86.31	3.56	0.10
CNN	**96.30**	0.76	90.93	4.37	7.33
DANN	95.17	0.92	91.58	3.01	6.25
MCD	95.29	0.46	**93.60**	**2.36**	7.33

The MCD achieves higher classification accuracy (mean = 93.60%) and lower standard deviation (2.36%) than the other methods for classifying the locomotion intent of the target subjects (see Table I). The significant effects of using different methods are also described as a P matrix in Table II (P is the probability that the null hypothesis is true). There is a significant difference ($P < 0.02$) between using the deep neural network (CNN, DANN, MCD) and using the traditional algorithms (LDA, SVM, and ANN) for both source subjects and target subjects, which validates the feasibility of the present network. The disadvantage of the present network is that it takes a longer time (7.33 ms) than the traditional algorithms to classify a signal segment. Although by using the same network architecture, the MCD classifies the locomotion intent of the target subjects with a 2.67% higher mean and a 1.91% lower standard deviation of the classification accuracy than for CNN, but their difference is not significant ($P = 0.11$). Compared to the DANN, which is a representative domain adaptation method,

TABLE II

P MATRIX OF CLASSIFICATION ACCURACY IN TABLE I. THE VALUE IN THE MATRIX INDICATES THE *P* VALUE BETWEEN THE CLASSIFICATION ACCURACY USING THE METHOD IN THE ROW AND THAT USING THE METHOD IN THE COLUMN. THE LOWER LEFT CORNER AREA REPRESENTS THE *P* VALUE OF THE CLASSIFICATION ACCURACY FOR THE SOURCE SUBJECTS, AND THE UPPER RIGHT CORNER AREA REPRESENTS THAT FOR THE TARGET SUBJECTS

Methods	LDA	SVM	ANN	CNN	DANN	MCD
LDA	/	0.023	0.772	0.019	0.003	0.000
SVM	0.000	/	0.011	0.000	0.000	0.000
ANN	0.000	0.000	/	0.018	0.002	0.000
CNN	0.000	0.000	0.000	/	0.703	0.106
DANN	0.000	0.000	0.000	0.008	/	0.111
MCD	0.000	0.000	0.000	0.002	0.704	/

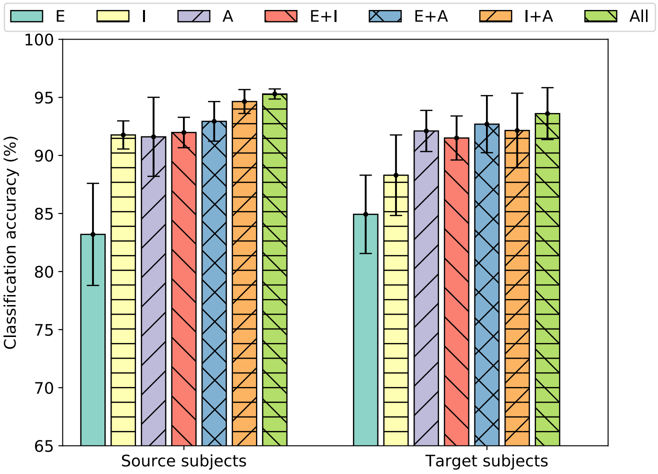


Fig. 4. Accuracy of classifying the locomotion intent for source subjects and the target subject on ENABL3S using the MCD method. The signals are divided into different groups based on the type of sensors: EMG (E), IMU (I), angle sensor (A), their combinations (+ denotes combination), and all sensors (All). The error bars represent mean ± one standard deviation of classification accuracy in the leave-one-subject-out test.

the MCD achieves a 2.02% higher mean and a 0.65% lower standard deviation of the classification accuracy on the target domain, and again their difference is not significant either ($P = 0.11$). Moreover, the MCD sacrifices its performance for classifying the locomotion intent of the source subjects. The mean accuracy of classifying the locomotion intent for the source subjects using the MCD is statistically lower than that using the CNN ($P = 2 \times 10^{-3}$). Therefore, the MCD outperforms the CNN in the target domain while achieving lower performance than the CNN in the source domain.

The effects of selecting different sensors were also analyzed to determine the importance of different types of sensors. The signals captured from different types of sensors were utilized to train the MCD. The P value of one-way ANOVA with post hoc test is 2.16×10^{-15} for the source subjects and 1.25×10^{-8} for the target subjects, which means that changing the sensor has a significant effect on classifying the human signals. As shown in Fig. 4, the EMG performs worse than the other sensors ($P < 3 \times 10^{-4}$ for the source subjects and $P \leq 0.05$ for the target subjects). This result makes sense because the EMG signals are noisy and highly user-dependent. Besides, there is no significant difference ($P = 0.89$) between the classification accuracy for the source subjects using the IMU and that using the angle sensors, but angle sensors significantly outperform the IMU in classifying the locomotion intent of the target subjects ($P < 0.01$). Hence, the angle sensors are less user-dependent than the IMUs and EMGs. The reason may be that the signals of joint angles are more stable than the signals of linear acceleration, angular velocities, and muscle signals. In addition, the human lower limb model can be built based on the joint angles, which may also explain the improved performance of the angle sensors. The mean values of the classification accuracy for the source subjects and the target subjects based on all sensors are 0.65% and 1.45% higher than those based on IMUs and angle sensors (I+A), but the differences are not significant ($P = 0.10$ for source subjects and $P = 0.28$ for target subjects). Hence, the EMG can still provide some supplementary information for predicting the human locomotion intent but need to be combined with kinetic sensors.

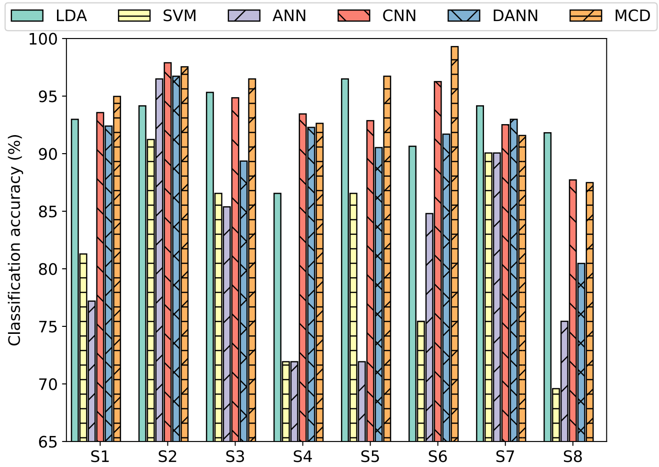


Fig. 5. Accuracy of classifying human activities for each target subject on DSADS using the LDA, SVM, ANN, CNN, DANN, and MCD methods. S1-S8 represent eight target subjects.

B. Evaluation on DSADS

The present method was also evaluated on the DSADS to classify human activities. The classification accuracy using LDA, CNN, DANN, and MCD is higher than that using SVM and ANN for most target subjects (see Fig. 5). Moreover, the range of classification accuracy reduces to about 11.81% after using LDA, CNN, and MCD, which is lower than that using the SVM, ANN, and DANN (≥ 16.26%).

The mean of classification accuracy (94.59%) for the target subjects using the MCD is still the highest (see Table III). However, CNN decreases the standard deviation of the classification accuracy to 3.01% for the target subjects. As shown in Table IV, the CNN and MCD statistically outperform the SVM and ANN in classifying the activities of the target subjects ($P < 3 \times 10^{-3}$). Although the MCD achieves 0.95% and 3.79% higher classification accuracy in the target domain than CNN and DANN, there is no significant difference between their classification accuracy ($P \geq 0.10$). It is surprising that

TABLE III

ACCURACY AND FORWARD TIME OF CLASSIFYING HUMAN ACTIVITIES FOR THE SOURCE SUBJECTS AND THE TARGET SUBJECT OF DSADS USING THE LDA, SVM, ANN, CNN, DANN, AND MCD METHODS. SD DENOTES THE STANDARD DEVIATION. THE FORWARD TIME INDICATES THE ONLINE EXECUTING TIME OF CLASSIFYING EACH SIGNAL SEGMENT. THE TIME IS IN MILLISECONDS (ms)

Methods	Mean (%)	SD (%)	Mean (%)	SD (%)	Time (ms)
	Source		Target		
LDA	98.02	**0.22**	92.76	3.12	0.07
SVM	96.78	1.15	81.58	8.36	**0.05**
ANN	97.72	0.27	81.65	8.96	0.13
CNN	**99.20**	0.73	93.64	**3.01**	7.72
DANN	98.55	0.87	90.80	4.70	6.82
MCD	97.73	0.81	**94.59**	3.83	7.72

TABLE IV

P MATRIX OF CLASSIFICATION ACCURACY IN TABLE III. THE MEANINGS OF THE VALUES ARE THE SAME AS THOSE IN TABLE II

Methods	LDA	SVM	ANN	CNN	DANN	MCD
LDA	/	0.003	0.005	0.576	0.343	0.313
SVM	0.010	/	0.987	0.002	0.017	0.001
ANN	0.032	0.041	/	0.003	0.023	0.002
CNN	0.001	0.000	0.000	/	0.172	0.590
DANN	0.115	0.004	0.022	0.128	/	0.099
MCD	0.361	0.077	0.973	0.002	0.074	/

there is no significant difference between the classification accuracy in the target domain using the LDA and that using CNN, DANN, and MCD ($P > 0.31$). This result shows that the LDA can classify IMU signals accurately. In addition, the LDA can optimize its parameters analytically and is time-efficient (forward time = 0.07 ms), and thus it is suitable for real-time application. In the source domain, all methods achieve high classification accuracy (≥ 96.78), which shows that the labeled training set is still very important. The CNN still significantly outperforms the other methods except the DANN (mean = 99.20% and $P \leq 2 \times 10^{-3}$), but the lowest standard deviation (0.22%) of the classification accuracy in the source domain is achieved by the LDA.

The subjects in DSADS wore five IMUs on different body parts, including torso (T), right arm (RA), left arm (LA), right leg (RL), and left leg (LL). The signals captured from different IMUs were utilized to train the MCD and test the classification accuracy for source subjects and target subjects. The classification accuracy was compared to evaluate the effects of sensor positions. Based on the result of the one-way ANOVA with post hoc test, different sensor positions do not cause significant difference for the classification accuracy in the source domain ($P = 0.37$) and the target domain ($P = 0.14$). After wearing all five IMUs, the mean of classification accuracy for the source subjects and the target subjects increase to 97.73% and 94.59%, respectively, which are 0.30% and 2.4% higher than the highest accuracy using the single IMU (see Fig. 6). However, the classification accuracy for using five IMUs does not significantly outperform that using the single IMU ($P > 0.07$) except that using the sensor on the left leg ($P < 8 \times 10^{-3}$).

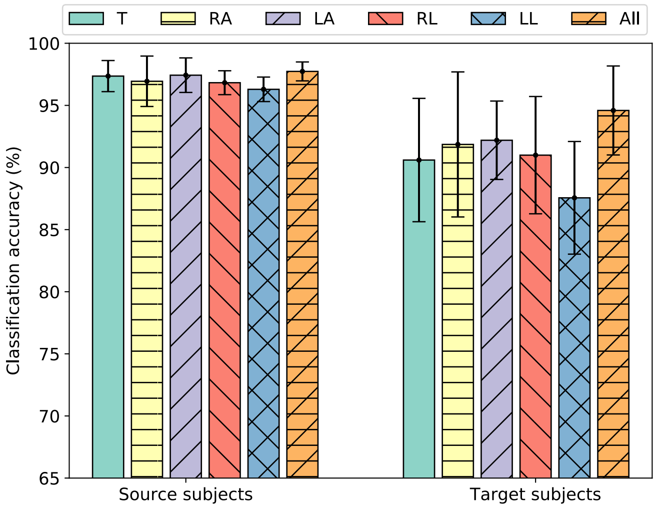


Fig. 6. Accuracy of classifying the human activities for source subjects and the target subject on DSADS using the MCD method. The signals are divided into different groups based on the position of IMU: torso (T), right arm (RA), left arm (LA), right leg (RL), and left leg (LL), and all IMUs (All). The error bars represent mean ± one standard deviation of classification accuracy in the leave-one-subject-out test.

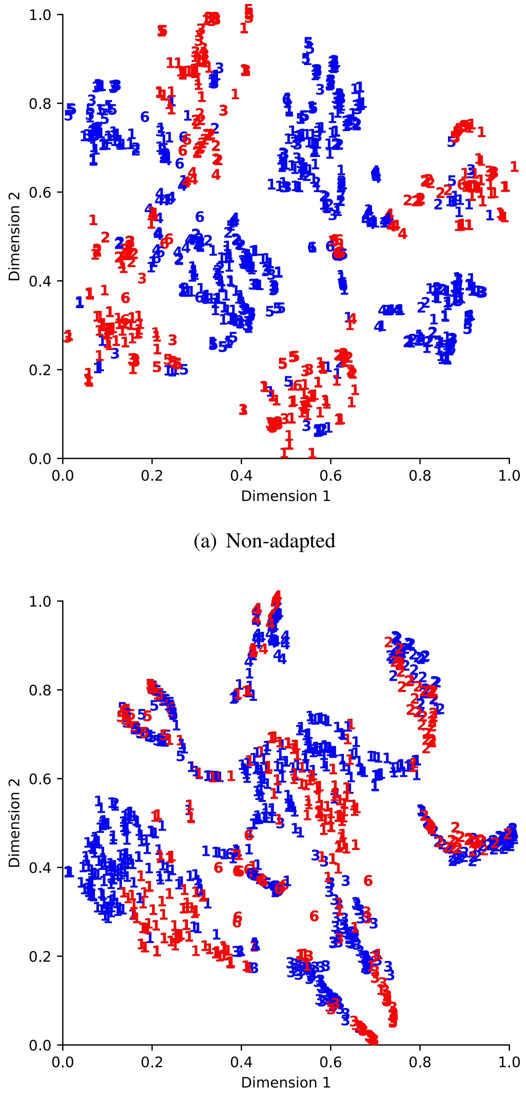
C. Visualization

To better visualize the different distributions of the hidden features between the source domain and the target domain, the present paper also visualizes t-SNE projection [35] of the non-adapted input features (Fig. 7(a)) and the adapted features (Fig. 7(b)) generated from the last layer of the feature generator of the MCD. The overlap between the different domains suggests the success of the adaptation and is positively related to the classification accuracy. As shown in Fig. 7, the features of different domains align better after using the MCD. Before adaptation, there is almost no overlap between the blue points (source domain) and the red points (target domain). After adaptation, the blue and the red points with the same labels distribute in a similar area.

IV. DISCUSSION

A. Ablation Study

In the present paper, an unsupervised cross-subject adaptation method (MCD) was presented to predict the locomotion intent and activities of humans. The MCD was trained by the labeled training set of the source subjects and the unlabeled training set of target subjects. Then the MCD was validated and tested on the validation set and testing set of the target subjects. Experimental results validated that the MCD achieved the highest accuracy (93.60% on ENABL3S and 94.59 on DSADS) to classify the locomotion intent and the activities of the target subjects. These results confirmed the feasibility of the unsupervised cross-subject adaptation method in classifying the locomotion intent and the activities of the target subjects, which could increase the user-independence of the classifiers and reduce the burden of labeling a large amount of data for each new subject. Besides these expected results, some other interesting results were found.



(a) Non-adapted

(b) Adapted

Fig. 7. (Best viewed in color) T-SNE visualization of the hidden features for one target subject from the ENABL3S dataset before and after adaptation using the MCD. Different colors represent different domains. Blue and red points represent the source and the target hidden features, respectively. The different numbers denote different classes of the hidden features. All input samples are from the testing set. After adaptation, the decision boundaries between each cluster are found to be better than those of the non-adapted case.

Firstly, the MCD and CNN, which shared the same network architecture, achieved different performances in the source domain and the target domain. The CNN performed better in the source domain (96.30% on ENABL3S and 99.20% on DSADS) while the MCD achieved higher accuracy in the target domain (93.60% on ENABL3S and 94.59% on DSADS). There was only one loss function for the CNN, which reflected the accuracy of classifying the data in the source domain. Conversely, there were three loss functions for the MCD. The first loss function was the same as that for CNN. The second loss function was used for training the classifiers to minimize the first loss function and maximize the discrepancy of the classifiers. The third loss function was used for training the feature generator to minimize the discrepancy of classifiers. The MCD with the first loss, with the first and the second loss, and with all three losses achieved the target accuracy at 90.93%, 92.13%, and 93.60%, respectively on ENABL3S and at 93.64%, 93.49%, and 94.59%, respectively, on DSADS. Hence, different losses and training strategies caused different performances. The last two adversarial loss functions should be combined to inspire the network to learn domain-independent features. Besides, the last two losses influenced the training objective of the MCD and thus decreased its performance in the source domain. It is difficult to optimize multiple objectives simultaneously. When the labels are available, the supervised learning method is still the best choice to train CNN. When the labels are not available, the unsupervised domain adaptation methods can be helpful.

Second, kinetic sensors, such as IMU and angle sensors can be utilized to predict the locomotion intent of subjects. As shown in Fig. 4, using only IMU and angle sensors (I+A), the mean accuracy for the source subjects and the target subjects achieved 94.64% and 92.15%, respectively, which were not significantly different from those using all sensors ($P = 0.10$ and $P = 0.28$). This result may change the traditional view that the signals of kinetic sensors generate behind the motion and cannot be used to predict the intent of humans. There are still some feedforward signals in IMU and angle sensors, such as linear acceleration and angular velocity. Based on the experimental results of the present paper, these feedforward signals could be effectively used to classify the human intent. Moreover, the EMG signals could help to increase the classification accuracy further. After using all the sensors, the mean accuracy increased by 0.56% and 1.45% in the source domain and the target domain, respectively, but these improvements are not significant. The EMG is more significant in directly predicting the joint angles or joint torques [2]. However, the EMG signals are noisier and may change during walking [36]. For the classification task, the number of EMG sensors can be reduced to increase the robustness and make the subjects feel more comfortable.

Third, the CNN can accurately classify human activities based on the IMU signals (mean accuracy = 99.2%, see Table III). There were more activities (19 activities) in the DSADS, which were supposed to be more difficult to recognize than ENABL3S (6 activities). However, experimental results showed that all methods could accurately classify human activities (accuracy > 96.78% in the source domain and > 81.58 in the target domain). The LDA also achieved high performance in the target domain (92.76%). The reasons may be that the IMU signals were more stable and the time length of the data segment on DSADS was 5 s, which was much longer than for ENABL3S (300 ms). The long data segment cannot be used to predict the locomotion intents but can be used to label the historical activities in real-time, which may also increase the user-independence of wearable robots.

Finally, the data size can affect the result of the domain adaptation. In previous papers [26], [27], the size of the source data was usually larger than that of the target data. In the present leave-one-subject-out test, there are more source subjects than the target subject, which can be seen as the many-to-one transfer. The present paper also implemented a one-to-many transfer experiment, where there were only one source subject and multiple target subjects. In the one-to-many transfer experiments, the classification accuracy on the target domain decreased to 80.57% and 76.09% for the ENABL3S and DSADS datasets, respectively. This phenomenon is not occasional, and it has been proved why increasing the size of the source data helps improve the performance of the classifier on the target domain (see the Appendix for the detailed theoretical explanation).

B. Comparison With Existing Works

The present MCD performed better for classifying target data than the respective domain adaptation method: DANN. The MCD achieved 2.02% and 3.79% higher classification accuracy than the DANN on the target domain of the ENABL3S and DSADS, respectively. The MCD and DANN have the same feature extractor and label classifier but have different training strategies. The DANN only classified the domain of the input data and entirely aligned the feature space of the source and the target data without considering the decision boundaries. Because the data of different classes may distribute differently, aligning the overall feature space may negatively affect the alignment of the features that belong to the same class. Comparatively, the MCD seems to be better because it uses two label classifiers to estimate the decision boundaries of the source and the target data and a common feature generator to align the features based on the estimated decision boundaries.

Hu *et al.* [9] have achieved high accuracy on ENABL3S (mean accuracy = 98.57%) to recognize the locomotion intent of subjects using the LDA. However, they assumed that the locomotion mode in the last step was known and trained 20 classifiers based on different locomotion modes in the last step. Besides, they also restricted the transition between different locomotion modes. For instance, they hypothesized that the next mode of stair ascent could only be stair ascent or level ground walking. Hence, they have utilized much prior information to further increase the classification accuracy. Moreover, they trained the classifiers for each subject, which was highly user-dependent. In this study, the accuracy was estimated based on the current segment of the signals without considering the last locomotion mode. Also, the transition between different locomotion modes was not limited. The present paper only used one classifier to classify the locomotion intent of both source subjects and target subjects. Besides, the present paper recognized six different locomotion modes, which were more than five modes in [9]. Compared to the results in [9], our results were just the original results of classifiers without being fine-tuned. Therefore, the LDA in the present study only achieved an accuracy of 92.59% and 85.78% to classify the locomotion intent of the source subjects and the target subjects, respectively. Comparatively, the presented CNN and MCD increased the accuracy by 3.71% and 7.82%, respectively, in the source domain and the target domain, respectively, which indicated the advantages of the developed methods. After using some fine-tuning strategies similar to those in [9] or the decision fusion method [37], the classification accuracy could be increased further.

Fallahzadeh and Ghasemzadeh [26] also presented an unsupervised domain adaptation method to classify human activities of the target subjects on DSADS and achieved 87% mean accuracy in the leave-one-subject-out test using the IMU placed on the torso. Using one IMU placed on the torso, the presented method achieved 90.60% mean accuracy for the target subjects in the leave-one-subject-out test (see Fig. 6). Moreover, the present paper also compared the results using a single IMU and all IMUs, and experimental results showed that the mean accuracy in the target domain could increase to 94.59% after using all IMUs. Therefore, it is important to wear multiple sensors and use sensor fusion methods to provide more stable information.

C. Limitations and Future Works

Although the present study fulfilled the goal of classifying the locomotion intent and activities of the target subjects, there are still some limitations. Firstly, the classification accuracy of the locomotion intent in the target domain should be increased further. Locomotion intent prediction is related to the real-time control of wearable robots, such as exoskeleton and prosthesis, and the incorrect prediction may cause the user to tumble. To resolve this issue, we will combine the decision fusion method and the present unsupervised cross-subject adaptation to filter the decisions and increase the classification accuracy. In our previous work [37], we designed an analytical method based on the hidden Markov model to fuse the sequential decisions, and the processing time for each decision was 3 milliseconds. Then the overall processing time for the present MCD and the decision fusion method will be 11 milliseconds, which is adequately short enough for a high-level controller. Additionally, we will label the previous locomotion modes using the IMU and other mechanical signals. Then the system will be able to obtain a large amount of labeled data of target subjects and train the network with the supervised learning method.

Second, the present method has only been evaluated on the dataset of subjects without disabilities, because we have not found a public dataset that includes the signals of subjects with disabilities who walk on different terrains. The gait patterns of subjects with disabilities are different from those of subjects without disabilities, which may decrease the classification accuracy for the target subject. If some labeled signals of subjects with disabilities are also included in the source dataset, the classification accuracy for new target subjects with disabilities may still be high because the present method is able to learn the common features of the subjects with disabilities and subjects without disabilities. To validate this assumption, we will capture and label signals of subjects with disabilities to prepare a more complete dataset and analyze the corresponding results in the future.

Thirdly, there are still some intents that cannot be predicted using the present method, for example, crossing an obstacle, turning around, and kicking a football. This is also a limitation of the finite-state controller because it is impossible to list all possible intents for a new user. We may also combine a finite-state controller with a volitional controller, which allows the subjects to directly control the joint angles of wearable robots using the muscle signals. However, volitional control is not robust. There are still many challenges for accurately predicting human intents.

Finally, the present study only performed offline analysis, and there may be some other challenges in an online test. Therefore, the presented method will be applied to the real-time control of the wearable robots to evaluate its performance in real-time.

V. Conclusion

Labeling a large amount of user-dependent human signals to accurately classify the locomotion intent of subjects has been a burden to both the researcher and the subject. The present paper resolved this problem by developing an unsupervised cross-subject adaptation method to accurately recognize the locomotion intent and the activities of target subjects. The developed method only utilized the labeled training data of source subjects and unlabeled training data of target subjects, which avoided the burdensome works of labeling data. The proposed method was evaluated using two public datasets (ENABL3S and DSADS), which recorded signals of able-bodied subjects while performing different activities. Experimental results showed that the presented method was able to classify the locomotion intent and activities of target subjects at high accuracy (93.60% on ENABL3S and 94.59% on DSADS), which validated the feasibility and the level of accuracy of the developed method. This study indicated that after capturing human signals and designing the unsupervised cross-subject domain adaptation method, robots can accurately predict locomotion intent of humans, which is beneficial for the control of wearable robots and for the improvement of human-robot interaction.

Appendix

Lemma 1: Let D_{SS} and D_{SL} be the source domain with a small distribution range and that with a large distribution range in the feature space, where $D_{SS} \subset D_{SL}$, and let D_T be the target domain. If data in D_{SS}, D_{SL}, and D_T can be classified correctly using the same idea classifier h^* trained by the supervised learning method, then for the classifier $h_{SS} \in H_{SS}$ and $h_{SL} \in H_{SL}$:

$$\sup_{h_{SS}\in H_{SS}} \epsilon_T^{SS}(h_{SS}) \ge \sup_{h_{SL}\in H_{SL}} \epsilon_T^{SL}(h_{SL}), \tag{6}$$

here H_{SS} and H_{SL} are two classifier spaces where $\epsilon_{SS}(h_{SS}) \to 0$ and $\epsilon_{SL}(h_{SL}) \to 0$; $\epsilon_{SS}(h_{SS})$ and $\epsilon_{SL}(h_{SL})$ denote the error in the source domain D_{SS} and D_{SL}, which can be minimized by the supervised learning with the labeled source data; $\sup_{h_{SS}\in H_{SS}} \epsilon_T^{SS}(h_{SS})$ and $\sup_{h_{SL}\in H_{SL}} \epsilon_T^{SL}(h_{SL})$ indicate the supremum of errors of the classifier h_{SS} and h_{SL} on the unlabeled target domain D_T transferred from the labeled source domain D_{SS} and D_{SL}, respectively.

Proof: According to previous research [34]:

$$\epsilon_T(h) \le \epsilon_S(h) + \frac{d_{H\Delta \mathrm{H}}(D_S, D_T)}{2} + \lambda = \sup_{h\in H} \epsilon_T^S(h), \tag{7}$$

where the $\epsilon_T(h)$ and $\epsilon_S(h)$ indicate the error of a classifier $h \in H$ on the target domain D_T and the source domain D_S, respectively. The H-distance for domain divergence is denoted by $d_{H\Delta \mathrm{H}}(D_S, D_T)$. The combined error of the ideal joint classifier h^* is indicated by λ:

$$\lambda = \min\left(\epsilon_S\left(h^*\right) + \epsilon_T\left(h^*\right)\right). \tag{8}$$

Since data in D_{SS}, D_{SL}, and D_T can be classified correctly using the same ideal classifier h^* trained by the supervised learning method, $\lambda \to 0$. Besides, the classifier h is trained by the labeled data in the source domain, $\epsilon_S(h) \to 0$. Therefore, the $\epsilon_T(h)$ mainly depends on $d_{H\Delta \mathrm{H}}(D_S, D_T)$. Based on the previous research [34], $d_{H\Delta \mathrm{H}}(D_S, D_T)$ is calculated as below:

$$\begin{aligned} d_{H\Delta \mathrm{H}}(D_S, D_T) = 2 \sup_{h,h'\in H} \Big| & Pr_{x\sim D_S}[h(x) \ne h'(x)] \\ & - Pr_{x\sim D_T}[h(x) \ne h'(x)]\Big|, \end{aligned} \tag{9}$$

where Pr denotes the probability. Two classifiers h and h' belong to the classifier space H.

Since two classifiers h and h' are both trained by the same labeled data in the source domain, their disagreement on the same data x in the source domain will be low: $Pr_{x\sim D_S}[h(x) \ne h'(x)] \to 0$. Therefore, $d_{H\Delta \mathrm{H}}(D_S, D_T)$ mainly depends on the supremum of the disagreement of two classifiers on the same data x in the target domain $\sup_{h,h'\in H}\left|Pr_{x\sim D_T}[h(x) \ne h'(x)]\right|$.

The sub classifier spaces H_{SS} and H_{SL} fulfill the requirement that $\epsilon_{SS}(h_{SS}) \to 0$, $h_{SS} \in H_{SS}$ and $\epsilon_{SL}(h_{SL}) \to 0$, $h_{SL} \in H_{SL}$. Recall that the distribution range of the source domain D_{SS} is smaller than that of the D_{SL}: $D_{SS} \subset D_{SL}$. The bigger the distribution range of a domain is, the smaller the sub classifier space is. The reason is that the larger distribution range introduces more constraints to classify the data in the domain. Therefore, $H_{SS} \supset H_{SL}$. Since the classifier space is the same for classifying the data in the source domain and the target domain, the supremum of the disagreement of two classifiers on the same data x in the target domain is also affected by different classifier spaces $H_{SS} \supset H_{SL}$:

$$\begin{aligned} & \sup_{h,h'\in H_{SS}} \left|Pr_{x\sim D_T}[h(x) \ne h'(x)]\right| \\ & \qquad \ge \sup_{h,h'\in H_{SL}} \left|Pr_{x\sim D_T}[h(x) \ne h'(x)]\right| + \delta, \end{aligned} \tag{10}$$

where δ is a small value. Then:

$$d_{H\Delta \mathrm{H}}(D_{SS}, D_T) \ge d_{H\Delta \mathrm{H}}(D_{SL}, D_T) + 2\delta. \tag{11}$$

Recall that $\epsilon_{SS}(h_{SS}) \to 0$, $h_{SS} \in H_{SS}$ and $\epsilon_{SL}(h_{SL}) \to 0$, $h_{SL} \in H_{SL}$. Hence, $|\epsilon_{SL}(h_{SL}) - \epsilon_{SS}(h_{SS})| \le \delta$ can be achieved using the supervised learning method with the labeled source data. Then:

$$\begin{aligned} \epsilon_{SL}(h_{SL}) &= \epsilon_{SS}(h_{SS}) + \epsilon_{SL}(h_{SL}) - \epsilon_{SS}(h_{SS}) \\ &\le \epsilon_{SS}(h_{SS}) + |\epsilon_{SL}(h_{SL}) - \epsilon_{SS}(h_{SS})| \\ &\le \epsilon_{SS}(h_{SS}) + \delta. \end{aligned} \tag{12}$$

Combine (11) with (12):

$$\begin{aligned}\epsilon_{SL}\left(h_{SL}\right) &+ \frac{d_{H\Delta \mathrm{H}}\left(D_{SL}, D_T\right)}{2} + \lambda \\ &\leq \epsilon_{SS}\left(h_{SS}\right) + \delta + \frac{d_{H\Delta \mathrm{H}}\left(D_{SS}, D_T\right)}{2} - \delta + \lambda \\ &= \epsilon_{SS}\left(h_{SS}\right) + \frac{d_{H\Delta \mathrm{H}}\left(D_{SS}, D_T\right)}{2} + \lambda. \qquad (13)\end{aligned}$$

Based on (7) and (13):

$$\sup_{h_{SS} \in H_{SS}} \epsilon_T^{SS}(h_{SS}) \geq \sup_{h_{SL} \in H_{SL}} \epsilon_T^{SL}(h_{SL}). \qquad (14)$$

References

[1] S.-J. Blakemore and J. Decety, "From the perception of action to the understanding of intention," *Nature Rev. Neurosci.*, vol. 2, no. 8, pp. 561–567, Aug. 2001.

[2] T. R. Clites *et al.*, "Proprioception from a neurally controlled lower-extremity prosthesis," *Sci. Transl. Med.*, vol. 10, no. 443, May 2018, Art. no. eaap8373.

[3] F. Sup, A. Bohara, and M. Goldfarb, "Design and control of a powered transfemoral prosthesis," *Int. J. Robot. Res.*, vol. 27, no. 2, pp. 263–273, Feb. 2008.

[4] M. Hao, J. Zhang, K. Chen, and C. Fu, "Design and basic control of extra robotic legs for dynamic walking assistance," in *Proc. IEEE Int. Conf. Adv. Robot. Social Impacts (ARSO)*, Oct. 2019, pp. 246–250.

[5] A. J. Young and D. P. Ferris, "State of the art and future directions for lower limb robotic exoskeletons," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 2, pp. 171–182, Feb. 2017.

[6] C. W. de Silva, *Sensors and Actuators: Engineering System Instrumentation*, 2nd ed. Boca Raton, FL, USA: CRC Press, Jul. 2015.

[7] M. R. Tucker *et al.*, "Control strategies for active lower extremity prosthetics and orthotics: A review," *J. Neuroeng. Rehabil.*, vol. 12, no. 1, pp. 1–29, Dec. 2015.

[8] R. Stolyarov, G. Burnett, and H. Herr, "Translational motion tracking of leg joints for enhanced prediction of walking tasks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 4, pp. 763–769, Apr. 2018.

[9] B. Hu, E. Rouse, and L. Hargrove, "Fusion of bilateral lower-limb neuromechanical signals improves prediction of locomotor activities," *Frontiers Robot. AI*, vol. 5, no. 78, pp. 1–16, Jun. 2018.

[10] H. L. Bartlett and M. Goldfarb, "A phase variable approach for IMU-based locomotion activity recognition," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 6, pp. 1330–1338, Jun. 2018.

[11] D. Xu, Y. Feng, J. Mai, and Q. Wang, "Real-time on-board recognition of continuous locomotion modes for amputees with robotic transtibial prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 10, pp. 2015–2025, Oct. 2018.

[12] M. Hao, K. Chen, and C. Fu, "Smoother-based 3-D foot trajectory estimation using inertial sensors," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 12, pp. 3534–3542, Dec. 2019.

[13] Z. Wu, J. Zhang, K. Chen, and C. Fu, "Yoga posture recognition and quantitative evaluation with wearable sensors based on two-stage classifier and prior Bayesian network," *Sensors*, vol. 19, no. 23, p. 5129, Nov. 2019.

[14] H. Huang, T. Kuiken, and R. Lipschutz, "A strategy for identifying locomotion modes using surface electromyography," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 1, pp. 65–73, Jan. 2009.

[15] L. Ambrozic *et al.*, "CYBERLEGs: A user-oriented robotic transfemoral prosthesis with whole-body awareness control," *IEEE Robot. Automat. Mag.*, vol. 21, no. 4, pp. 82–93, Dec. 2014.

[16] E. Zheng and Q. Wang, "Noncontact capacitive sensing-based locomotion transition recognition for amputees with robotic transtibial prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 2, pp. 161–170, Feb. 2017.

[17] C. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). New York, NY, USA: Springer-Verlag, 2006.

[18] M. Islam and E. T. Hsiao-Wecksler, "Detection of gait modes using an artificial neural network during walking with a powered ankle-foot orthosis," *J. Biophys.*, vol. 2016, pp. 1–9, Nov. 2016.

[19] K. Zhang, C. W. de Silva, and C. Fu, "Sensor fusion for predictive control of human-prosthesis-environment dynamics in assistive walking: A survey," Mar. 2019, *arXiv:1903.07674*. [Online]. Available: https://arxiv.org/abs/1903.07674

[20] O. Dehzangi, M. Taherisadr, and R. Changalvala, "IMU-based gait recognition using convolutional neural networks and multi-sensor fusion," *Sensors*, vol. 17, no. 12, p. 2735, Nov. 2017.

[21] J. Wang, V. W. Zheng, Y. Chen, and M. Huang, "Deep transfer learning for cross-domain activity recognition," in *Proc. 3rd Int. Conf. Crowd Sci. Eng. (ICCSE)*, Singapore, 2018, pp. 1–8.

[22] J. Wang and K. Zhang, "Unsupervised domain adaptation learning algorithm for RGB-D staircase recognition," Mar. 2019, *arXiv:1903.01212*. [Online]. Available: https://arxiv.org/abs/1903.01212

[23] Y. Massalin, M. Abdrakhmanova, and H. A. Varol, "User-independent intent recognition for lower limb prostheses using depth sensing," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 8, pp. 1759–1770, Aug. 2018.

[24] C. Liu, L. Song, J. Zhang, K. Chen, and J. Xu, "Self-supervised learning for specified latent representation," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 1, pp. 47–59, Jan. 2020.

[25] K. Zhang *et al.*, "Environmental features recognition for lower limb prostheses toward predictive walking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 465–476, Mar. 2019.

[26] R. Fallahzadeh and H. Ghasemzadeh, "Personalization without user interruption: Boosting activity recognition in new subjects using unlabeled data," in *Proc. 8th Int. Conf. Cyber Phys. Syst. (ICCPS)*, Pittsburgh, PA, USA, 2017, pp. 293–302.

[27] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 59, pp. 1–35, 2016.

[28] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 343–351.

[29] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971.

[30] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 3723–3732.

[31] B. Hu, E. Rouse, and L. Hargrove, "Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals," *Frontiers Robot. AI*, vol. 5, no. 14, pp. 1–5, 2018.

[32] B. Barshan and M. C. Yuksek, "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units," *Comput. J.*, vol. 57, no. 11, pp. 1649–1667, Nov. 2014.

[33] H. Huang, F. Zhang, L. J. Hargrove, Z. Dou, D. R. Rogers, and K. B. Englehart, "Continuous locomotion-mode identification for prosthetic legs based on neuromuscular-pmechanical fusion," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10, pp. 2867–2875, Oct. 2011.

[34] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, nos. 1–2, pp. 151–175, May 2010.

[35] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[36] B. H. Hu, E. J. Rouse, and L. J. Hargrove, "Using bilateral lower limb kinematic and myoelectric signals to predict locomotor activities: A pilot study," in *Proc. 8th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2017, pp. 98–101.

[37] K. Zhang, W. Zhang, W. Xiao, H. Liu, C. W. De Silva, and C. Fu, "Sequential decision fusion for environmental classification in assistive walking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 9, pp. 1780–1790, Sep. 2019.