Unsupervised Sim-to-Real Adaptation for Environmental Recognition in Assistive Walking

Chuheng Chen, Kuangen Zhang, Yuquan Leng, Xinxing Chen, *Member, IEEE*, and Chenglong Fu, *Member, IEEE*

***Abstract*— Powered lower-limb prostheses with vision sensors are expected to restore amputees' mobility in various environments with supervised learning-based environmental recognition. Due to the sim-to-real gap, such as real-world unstructured terrains and the perspective and performance limitations of vision sensor, simulated data cannot meet the requirement for supervised learning. To mitigate this gap, this paper presents an unsupervised sim-to-real adaptation method to accurately classify five common real-world (level ground, stair ascent, stair descent, ramp ascent and ramp descent) and assist amputee's terrain-adaptive locomotion. In this study, augmented simulated environments are generated from a virtual camera perspective to better simulate the real world. Then, unsupervised domain adaptation is incorporated to train the proposed adaptation network consisting of a feature extractor and two classifiers is trained on simulated data and unlabeled real-world data to minimize domain shift between source domain (simulation) and target domain (real world). To interpret the classification mechanism visually, essential features of different terrains extracted by the network are visualized. The classification results in walking experiments indicate that the average accuracy on eight subjects reaches (98.06% ± 0.71%) and (95.91% ± 1.09%) in indoor and outdoor environments respectively, which is close to the result of supervised learning using both type of labeled data (98.37% and 97.05%). The promising results demonstrate that the proposed method is expected to realize accurate real-world environmental classification and successful sim-to-real transfer.**

***Index Terms*— Unsupervised domain adaptation, lower-limb prostheses, sim-to-real transfer, environmental recognition, visualization.**

I. Introduction

LOWER-LIMB amputation attenuates the locomotion ability and physical function of millions of people [1], [2]. The emergence of lower-limb prostheses, especially powered prostheses, has made a great contribution to the recovery of amputees' mobility [3], including walking biomechanics and gait symmetry improvement along with falling risk and metabolic cost reduction [3]–[6]. However, when walking in complex environments, the amputee may be in an inconsistent locomotion mode with that of powered prosthesis, which may lead to serious consequence such as falling. Therefore, to better support amputees' terrain-adaptive locomotion, prostheses are required to predict the amputee's motion intent and switch to an appropriate locomotion mode accordingly in real time [7], [8].

Environmental information obtained by vision can guide not disabled individuals to switch their locomotion modes in response to changes in terrain [9]. For amputees with impaired vision-locomotion loops, vision sensors such as cameras can be used to obtain environmental information [10]–[12]. And the environment can be recognized with machine learning-based classification algorithms, providing environmental context for motion intent prediction [13]–[16]. Deep learning algorithms such as convolutional neural networks (CNNs) have achieved great success in image processing and classification [17], [18] and have also been increasingly used in recent environmental recognition studies [15], [19], [20]. In our previous researches [20], [21], a CNN was designed and trained with supervised learning and achieved an outstanding classification accuracy on environmental point cloud captured by a depth camera.

Nevertheless, supervised learning approaches also introduce a major challenge: the requirement of a large, well-labeled real-world dataset [22]. To improve the classification accuracy and the generalization ability in different environments,

Manuscript received January 10, 2022; revised April 5, 2022; accepted May 12, 2022. Date of publication May 18, 2022; date of current version May 27, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant U1913205 and Grant 62103180; in part by the Guangdong Innovative and Entrepreneurial Research Team Program under Grant 2016ZT06G587; in part by the China Postdoctoral Science Foundation under Grant 2021M701577; in part by the Science, Technology and Innovation Commission of Shenzhen Municipality under Grant SGLH20180619172011638, Grant ZDSYS20200811143601004, and Grant KYTDPT20181011104007; in part by the Stable Support Plan Program of Shenzhen Natural Science Fund under Grant 20200925174640002; and in part by the Centers for Mechanical Engineering Research and Education at Massachusetts Institute of Technology (MIT) and Southern University of Science and Technology (SUSTech). *(Chuheng Chen and Kuangen Zhang contributed equally to this work.) (Corresponding author: Chenglong Fu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Southern University of Science and Technology (SUSTech) Medical Ethics Committee under Approval No. 20210009.

Chuheng Chen, Yuquan Leng, Xinxing Chen, and Chenglong Fu are with the Shenzhen Key Laboratory of Biomimetic Robotics and Intelligent Systems and the Guangdong Provincial Key Laboratory of Human-Augmentation and Rehabilitation Robotics in Universities, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: fucl@sustech.edu.cn).

Kuangen Zhang is with the Shenzhen Key Laboratory of Biomimetic Robotics and Intelligent Systems and the Guangdong Provincial Key Laboratory of Human-Augmentation and Rehabilitation Robotics in Universities, 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.

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2022.3176410, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2022.3176410

researchers need to collect and label significant and diverse environmental data. In some recent studies [19], [23], [24], the images to be labeled may reach hundreds of thousands, which is time-consuming and laborious [25]. An attractive alternative to address such limitations is to automatically generate a large labeled synthetic dataset by simulation [20], [26], [27]. But with this comes another problem that simulated data presents some differences from real-world data regarding feature distribution, leading to inferior real-world classification performance for a model trained purely with it. These differences are collectively referred to as the simulation-to-real-world (sim-to-real) gap. It consists of real-world unstructured terrain and light conditions, the perspective and performance limitations of vision sensors, and other factors difficult to simulate [28], [29].

To mitigate this gap and thus realize sim-to-real transfer, unsupervised domain adaptation (UDA) has been used to better generalize models trained in a label-rich source domain (simulation) to a label-free target domain (real world) [22], [27], [30]. Many adversarial-based UDA methods such as domain adversarial neural networks (DANN) [22], [31] and adversarial discriminative domain adaptation (ADDA) [32] attempt to train a discriminator and different number of feature generators. The discriminator distinguishes the domain of hidden features. And the generators are trained to realize distribution alignment between the source features with that of the target to fool the discriminator. However, these approaches only attempt to make their distributions similar without considering the classification task of target samples. Thus a trained generator may generate ambiguous features near the decision boundary. In addition, to the authors' knowledge, there are no existing sim-to-real studies focusing on environmental recognition for wearable robots and generating simulated data for them. When the simulation is significantly different from the real world, forcefully aligning domains may lead to domain misalignment or even negative transfer [33], posing risks for amputees' locomotion.

To address these limitations and better overcome the sim-to-real gap, an unsupervised sim-to-real adaptation approach is proposed in this paper to accurately classify the real-world environments, providing low cost and robust support for motion intent prediction. It is hypothesized that the model trained with simulated data and unlabeled real-world data can still achieve precise real-world environmental classification. To that end, augmented simulated data with random dimensional and geometric noise are generated from a virtual camera perspective to reduce domain shift. Meanwhile, inspired by Saito *et al.*. [34], an end-to-end UDA method is incorporated to train the designed neural network consisting of a feature extractor and two different classifiers to align hidden features of simulated and real environments. This alignment is performed in an adversarial manner by training the classifiers to maximize their classification discrepancy in target domain, and then training the extractor to minimize the discrepancy. The real-world classification performance of this method is verified by indoor and outdoor experiments for able-bodied subjects and amputees. To visually explain the classification mechanism of the network, essential features of different terrains extracted by the network are clarified by visualization of class activation map.

The primary contributions of the present paper include:

1) Developing an unsupervised sim-to-real adaptation method to accurately classify real environments, providing environmental context for motion intent prediction.
2) Mitigating sim-to-real gap by generating randomized simulated environments from a camera perspective and training the designed network to realize feature alignment between simulated and real environments.
3) Visualizing the essential features of different terrains and presenting visual interpretation of the classification decision of our network.
4) Evaluating the performance of the proposed sim-to-real approach by indoor and outdoor walking experiments for able-bodied subjects and amputees.

The rest of the paper is organized as follows. Section II describes the proposed unsupervised sim-to-real adaptation approach. The experiment and visualization results are stated in Section III and discussed in Section IV. The conclusion of this paper is presented in Section V.

II. Methods

The proposed sim-to-real adaptation method is described in this section. An overview of this method is shown in Fig. 1. To maintain the integrity, the acquisition and preprocessing of real-world environmental data is briefly introduced, which has been thoroughly stated in our previous work [20]. Then, the generation of augmented simulated data, the unsupervised domain adaptation method and visualization method are presented. To verify the effectiveness of the proposed method, indoor and outdoor walking experiments are conducted.

A. Real-World Environmental Data Preprocessing

The 3D environmental point cloud captured by the depth camera has some limitations. First, in each gait cycle, the coordinate system of the camera fixed on the user's leg changes substantially with leg rotation. Thus, the output point cloud will also change significantly. To solve this problem and help distinguish level ground and ramp, rotation offset of the point cloud can be realized by switching the reference coordinate system of the point cloud to the invariant ground system in real-time with the help of the inertial measurement unit (IMU):

$$\boldsymbol{p}^{\text{Ground}} = \boldsymbol{R}^{\text{Ground}}_{\text{Camera}} \boldsymbol{p}^{\text{Camera}}, \tag{1}$$

where $\boldsymbol{p}^{\text{Ground}}$ and $\boldsymbol{p}^{\text{Camera}}$ are the point clouds in the ground and camera coordinate systems respectively; $\boldsymbol{R}^{\text{Ground}}_{\text{Camera}}$ denotes the rotation matrix from the camera system to the ground system, which is calculated based on the Euler angle of IMU.

Another problem is that the original 3D point cloud is unstructured and disordered. To avoid the time-consuming recognition of the 3D point cloud, a feasible solution is to project the point cloud into the sagittal plane (i.e., x-o-z plane) and convert it into binary image (see Fig. 2). As stated in our previous work [19], the 2D projection of the point cloud has

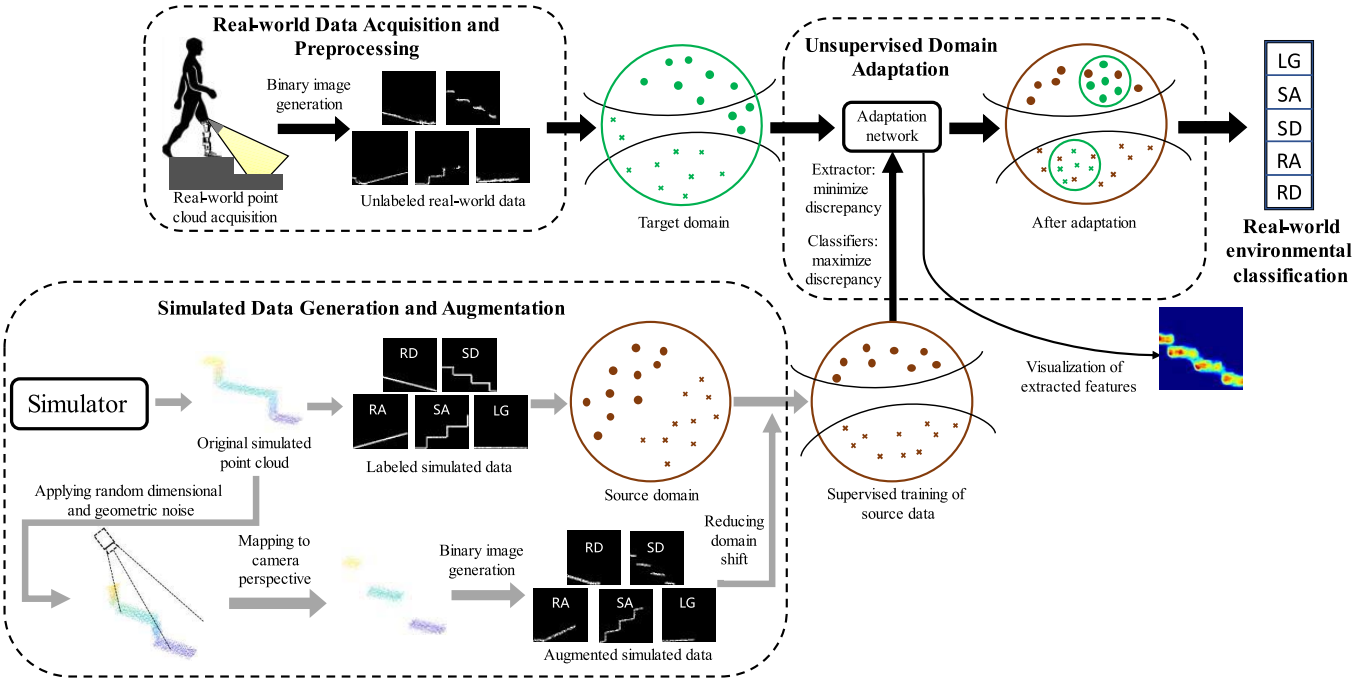


Fig. 1. The overview of the proposed unsupervised sim-to-real adaptation approach. Supervised learning on simulated data presents difficulty in performing real-world classification tasks. To address this, augmented simulated data with random dimensional and geometric noise is generated from camera perspective and supplemented to simulation dataset to augment it with more variability and reduce domain shift. The hidden features are aligned by training an adaptation network by unsupervised domain adaptation with simulated data and unlabeled real-world data. The features extracted from different terrains are visualized to make a visual interpretation of the classification mechanism. LG, SA, SD, RA and RD represent level ground, stair ascent, stair decent, ramp ascent, and ramp descent, respectively.

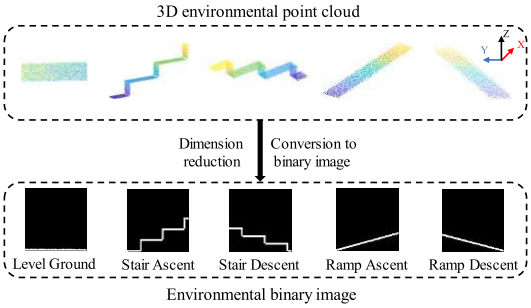


Fig. 2. Generation of environmental binary images. The 3D point cloud on the first row is a set of 3D points. The point cloud is projected to sagittal plane and converted to binary images (the second row).

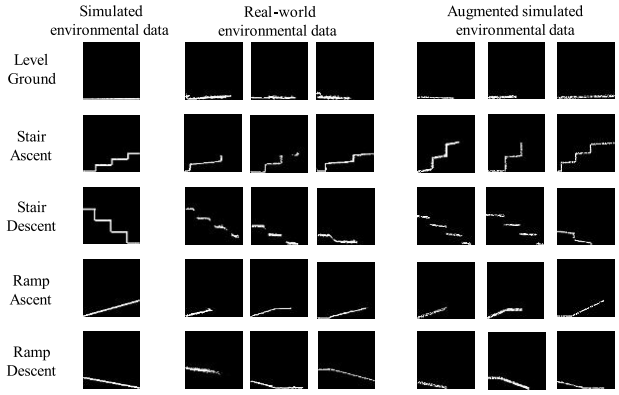


Fig. 3. Simulated and real-world data of different terrains. The left column is simulated data from the third-person perspective, which presents reality gaps from the real-world data. This gap is reduced by generating augmented simulated data with random dimensional and geometric noise from a virtual camera perspective. The real-world data is used for preliminary evaluation of the augmentation of simulated data.

enough information for environmental classification. The point cloud can be converted to a binary image as:

$$\boldsymbol{img}(r,c) = \begin{cases} 1 & 0.01\,(c-1) \le x_i - x_{\min} < 0.01c \;\&\\ & 0.01(r-1) \le z_i - z_{\min} < 0.01r \\ & i = 1, 2, \ldots\ldots, n \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $\boldsymbol{img}\ (r,\ c)$ represents the pixel in row r and column c of the image and each pixel corresponds to 0.01 m in the real world; x_i and z_i are the x and z coordinates of the i-th point in the 2D point cloud composed of n points; $x_{\min}$ and $z_{\min}$ represent the minimum x and z coordinates of the point cloud respectively.

B. Simulated Data Generation and Augmentation

In our previous research [20], 3D simulated environmental point cloud of five common terrains was generated according to their characteristics, including level ground, stair ascent, stair descent, ramp upsent and ramp descent. The geometries of different terrains were detailed in [20]. The generated point cloud was then converted into binary images based on the method in Section II.A. However, real-world environmental data presents reality gaps from the generated simulated data, which includes the following dimensions (see Fig. 3):

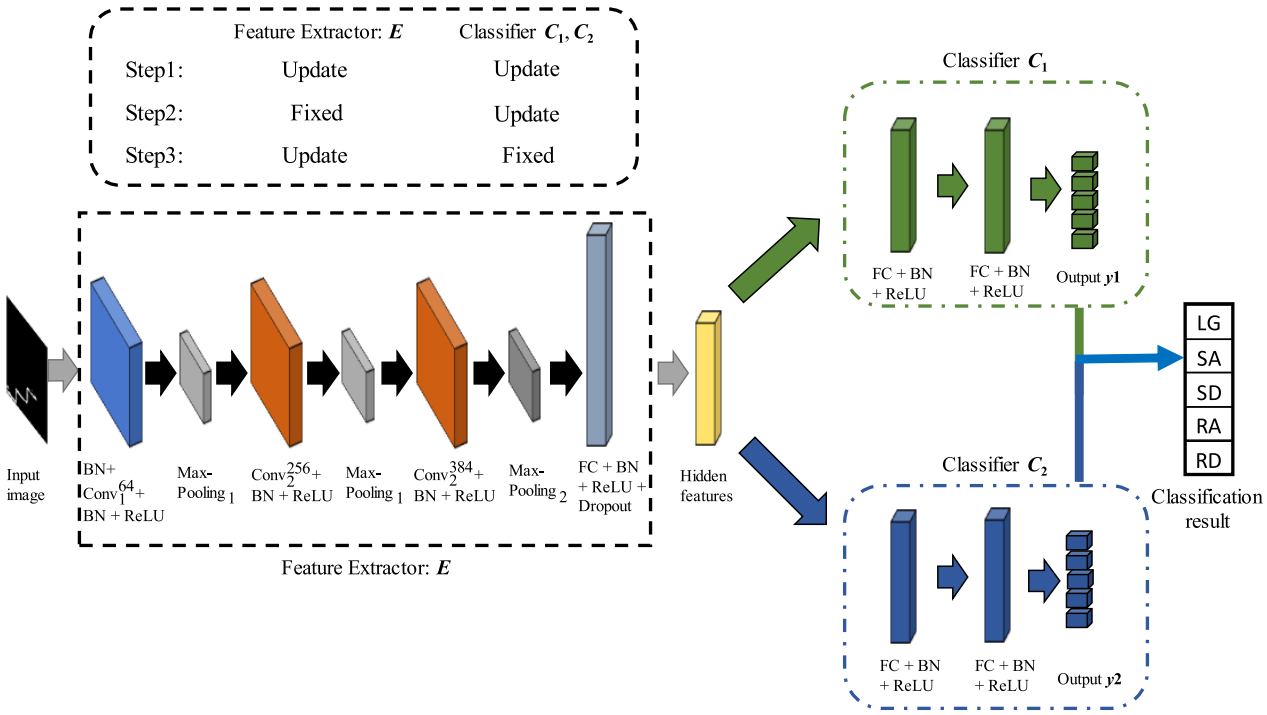


Fig. 4. The architecture and training steps of the proposed adaptation network. The type and parameter of a convolutional layer is indicated by both of the superscript and the subscript, e.g., Conv_2^{256} suggests that a type-two convolutional layer with 256 channels. The batch normalization layer and fully connected layer is denoted by BN and FC respectively. Training steps and objective functions are described in Section II.C.2) and 3-5.

1) There exists unstructured terrain such as the transition state between different terrains and the variance of the dimensions (e.g., length and height) and geometry (e.g., flatness and angularity) of each kind of terrain in real-world environmental data, especially in outdoor data.
2) Some environmental information is lost due to the camera perspective limitation such as the information about the vertical plane of stair descent, which differs from the intact simulated data from the third-person perspective.
3) Real-world point cloud is sparser because of light conditions and camera performance limitations.

Due to the aforementioned gaps, simulated data cannot recreate the diversity and noise in their real-world counterparts. Thus, the real-world classification performance of the CNN trained only with these data is poor. In this paper, to reduce the sim-to-real gap at the input level, augmented simulated environmental point cloud closer to the real world is generated (see Fig. 3), reducing this gap in the following aspects:

1) Applying random dimensional (the noise of height is only applied for each step of stairs) and geometric noise to all kinds of terrains and adding transition states between level ground and other terrains to simulate real-world.
2) Mapping the generated environmental point cloud to the perspective of a virtual camera fixed on the leg to simulate the perspective of a real-world depth camera.
3) Removing randomly scaled points in the generated point cloud to simulate the sparsity of real-world point cloud.

The augmented simulated point cloud is then converted into binary images and added to simulation dataset to augment the dataset with more diversity and enhance the adaptation performance. However, considering that some other reality factors affecting the classification performance such as various road conditions and ground objects are difficult to simulate, real environmental data and domain adaptation are still necessary for accurate real-world environmental classification.

C. Unsupervised Domain Adaptation

The problem of unsupervised domain adaptation (UDA) is defined as follows. Given a labeled source dataset $\{X_s, Y_s\}$ of source images (x_s, y_s) (where y_s is the label of x_s) and an unlabeled target dataset X_t of target images x_t, the source dataset from the source domain D_s is sufficient to train a model to perform the source classification task T_s. However, the shifted domain distribution of the target domain D_t from D_s indicates that the model trained for T_s cannot directly perform the target classification task T_t. Therefore, the aim of UDA is to perform T_t successfully by feature alignment between the source domain and the target domain.

1) Network Structure: Inspired by a theorem proposed by Ben-David *et al.*. [35] on how to make classifiers trained on source data perform well on target data, a neural network is designed for domain adaptation. It consists of a feature extractor $\boldsymbol{E}$ and two different classifiers $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$. The feature extracted by $\boldsymbol{E}$ from the input environmental data are shared by $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$. The final result is the category corresponding to the maximum value of the summed and normalized classification scores output by the two classifiers over the 5 types of terrains. The overall structure of the network and the training strategies are shown in Fig. 4.

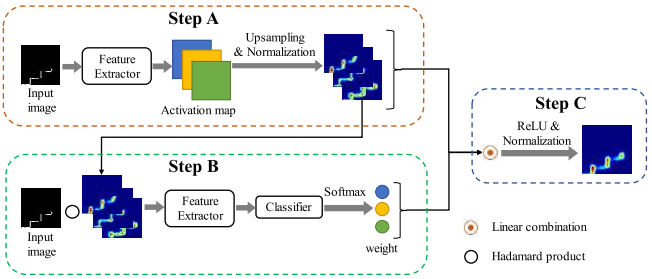


Fig. 5. The overview of the implemented CAM visualization method. The visualization result is obtained by linearly weighted summation of the extracted activation map in **step A** and the response value in **step B** as weight.

a) Feature extractor: The feature extractor is designed based on CNN because of its ability to automatically extract deep features and the parameter sharing method of convolutional layers, which can help the network learn efficiently.

The feature extractor consists of three convolutional layers, one fully-connected (FC) layer, three max-pooling layers, four activation layers with Rectified Linear Unit (ReLU), and one dropout layer. The two types of convolutional layers (Conv_1 and Conv_2) in the extractor consist of 5×5 filters with a stride of 1 and 3×3 filters with a stride of 1, respectively. Batch normalization (BN) is applied after each convolutional layer and FC layer to improve the generalization capability. Two types of max-pooling layer (MaxPooling_1 and MaxPooling_2) for downsampling have a kernel size of 7 and a stride of 2 and a kernel size of 3 and a stride of 1 respectively. The hidden features output by the last max-pooling layer are flattened and fed into an FC layer to reduce the feature dimension, followed by a dropout layer to avoid over-fitting.

b) Classifiers: The two classifiers share the same structure of three FC layers, two ReLU layers, and two BN layers. These FC layers map the features from the extractor to classification scores of 5 different environment classes.

2) Training Procedure: UDA trains the designed neural network for feature alignment between the source and target domains while considering the relationship between target samples and decision boundaries. Inspired by Saito *et al.*. [34], this objective is achieved through the following algorithm with three steps.

Step 1: First, the feature extractor and classifier are trained to correctly classify the source data. Before training, $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ need to be initialized separately to obtain different parameters. The training objective is to ensure our network can correctly classify different environmental data by minimizing the softmax cross-entropy loss on the source domain:

$$\min_{E,C_1,C_2} [-\mathbb{E}_{(x_s,y_s)\in(X_s,Y_s)} \sum_{k=1}^{K} I[k = y_s] \log \boldsymbol{P}_k(y|x_s)], \quad (3)$$

where $\boldsymbol{P}_k$ indicates the output probability of class k given the source input x_s; $I[k = y_s]$ is a binary indicator which equals 1 when k equals y_s and equals 0 otherwise; K indicates the number of classes of environment; $\mathbb{E}$ is an expectation operator.

Step 2: The second step requires fixing the parameters of $\boldsymbol{E}$ while training $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ to maximize the discrepancy of their output in the target domain. The different classification results of the two classifiers for a target sample mean that the sample is outside the source support. Such samples are prone to be misclassified, and also represent the discrepancy between the source and target domains. The objective is to enable the classifier to detect as many these target samples as possible while maintaining accurate classification of source samples:

$$\min_{C_1,C_2} [-\mathbb{E}_{(x_s,y_s)\in(X_s,Y_s)} \sum_{k=1}^{K} I[k = y_s] \log \boldsymbol{P}_k(y|x_s) - \mathbb{E}_{x_t\in X_t} \frac{1}{K} \sum_{k=1}^{K} (\left| \boldsymbol{P}_k^1(y|x_t) - \boldsymbol{P}_k^2(y|x_t) \right|)], \quad (4)$$

where $\boldsymbol{P}_k^1$ and $\boldsymbol{P}_k^2$ indicate the output probability for class k by $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$, respectively.

Step 3: Finally, while fixing the parameters of $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$, $\boldsymbol{E}$ is trained to minimize the discrepancy between classifiers on target samples. This makes the extracted target features gradually closer to source domain, and finally minimize the error on target domain and realize feature alignment. Besides, this step is repeated four times for each mini-batch to better update extractor parameters. The objective is defined as:

$$\min_{E} [\mathbb{E}_{x_t\in X_t} \frac{1}{K} \sum_{k=1}^{K} (\left| \boldsymbol{P}_k^1(y|x_t) - \boldsymbol{P}_k^2(y|x_t) \right|)]. \quad (5)$$

By repeating these steps, feature alignment between the source and target domains is realized, which enables the trained network to accurately perform the target classification task, i.e., real-world environmental classification.

3) Implementation Details: The proposed network is designed and trained on Python 3.7, CUDA 10.1 and PyTorch 1.8.0. In the training process, the Adam algorithm is used as the optimizer with a learning rate of 0.0002 and a weight decay of 0.0005. The maximum epoch is 100 and the batch size is 64. The training process and the experimental analysis are carried out on a computer equipped with an AMD Ryzen 7 4800H CPU (2.9 GHz), 16 GB RAM and an NVIDIA Geforce RTX 2060 GPU.

D. Visualization

Although CNN has achieved outstanding performance in various tasks in vision fields such as classification and detection, the logic of CNN decisions remains unclear. Visualization can be helpful to reason this logic and explain it in a way that people can understand. Inspired by Wang *et al.* [36], we employ a visualization method based on Class Activation Map (CAM) which provides visual interpretation with a weighted sum of activation maps from a chosen layer [37]. CAM clarifies the essential features extracted from different terrains and thus explains the classification decisions of our network. The overall process of the visualization consists of three steps (see Fig. 5).

Step A: The activation map A extracted from a specific layer of the feature extractor is upsampled and normalized as:

$$H^n = norm(Up(A^n)), \quad (6)$$

where H^n denotes the upsampled and normalized activation map; A^n is the n-th channel of A; $Up(\cdot)$ and $norm(\cdot)$ indicate the operations that upsample the activation map into the input size and normalize each element in the input to [0, 1] respectively.

Step B: The normalized feature map is used as a mask in the second step, and its Hadamard product with the original input image is fed into the network to obtain the response value of the image on the target category, which can be defined as:

$$\alpha_n^k = f(X \odot H^n) - f(X_b), \tag{7}$$

where α_n^k denotes the response value on category k; $f(\cdot)$ indicates the function of CNN; X is the input image; X_b is a known baseline input, which in this paper is set as a null matrix.

Step C: The final visualization result is the linearly weighted summation of the extracted activation map in **step A** and the response values obtained in **step B** as weight. A ReLU function was also applied on the weighted sum:

$$L_{CAM}^k = ReLU(\sum_k (\alpha_n^k A^n)). \tag{8}$$

Through the aforementioned steps, the essential features of different terrains obtained from the feature extractor are visualized, which helps to explain the classification mechanism and analyze the performance of our adaptation network.

E. Experimental Setup

To evaluate the proposed sim-to-real approach, six able-bodied subjects and three amputees were invited to participate in indoor and outdoor experiments to obtain environmental data. Each subject worn a depth camera and an IMU (MTi 1-series, Xsens, Netherlands) and walked repeatedly for five times under five kinds of indoor and outdoor terrain, including level ground, stair ascent, stair descent, ramp upsent and ramp descent. The comprehensive experimental setup is presented in [19].

The time-of-flight (ToF) depth camera (CamBoard pico flexx, pmdtechnologies, Germany) and the IMU were placed together on each subject's upper patellar tendon to capture environmental information in front of the subject. The capture rates of the environmental point cloud were 25 and 15 frames per second in indoor and outdoor experiments respectively. IMU calculated the Euler angles of the camera at a frequency of 100 Hz. Data from IMU and the camera were acquired in two threads, and their approximate synchronization was achieved by capturing and fusing the latest data from both threads.

After acquiring environmental data, the proposed network was trained using simulated data as source data and unlabeled real-world data as target data. Before training, the target dataset was randomly split into an training set (80%) and a manually labeled validation set (20%) for preliminary measurement of the performance and optimization of the hyper-parameters. To demonstrate the generalization capability of the network for different subjects, the target dataset only included the data of one able-bodied subject (Subject 0). After training, the data of other able-bodied subjects (Subject 1-5) and amputees (Amputee 1-3) are manually labeled and utilized as a test set to evaluate the indoor and outdoor classification performance of the network. To further enhance the accuracy, a mode filter was implemented to decrease the incidental error of the classification result. The filter smoothed the results through a sliding window, replacing the result of each frame with the mode of all classification result in this window. In this paper, the size of the window is set as 8.

In order to analyze the classification performance more comprehensively, this paper also trained a CNN by supervised learning with original simulated data, augmented simulated data, and a combination of simulated and real environmental data separately and compared their performance. The training parameters and the structure of the CNN were the same as those of the adaptation network, except that only a single classifier was used. In addition, the present paper also compared the performance of the proposed approach with DANN, which is considered as the baseline of the adaptation performance. The implemented DANN in the present paper used the same training strategy as proposed by Ganin *et al.* [31] and shared the same network structure as ours except for a domain classifier.

Finally, the visualization of CAM visually explained the classification decisions of the network. Similarly, the essential features extracted by the feature extractor trained only with the simulated data were visualized to verify the effect of adaptation on the extractor. In addition, to verify the effect of our approach on feature distribution, t-Distributed Stochastic Neighbor Embedding (t-SNE) projection [38] is used to visualize the features output by the last layer of the feature extractor.

F. Subject Information

Six able-bodied subjects (5 males and 1 female) and three amputees (all males) were invited to participate in our experiment. The able-bodied subjects were from the Southern University of Science and Technology, and the amputees were recruited from a local prosthetic company. The average age, height and body weight for able-bodied subjects and amputees were 26.5 $\pm$ 2.4 years old and 39.3 $\pm$ 1.6 years old, 168.5 $\pm$ 2.8 cm and 169.7 $\pm$ 0.4 cm, 59.3 $\pm$ 3.1 kg and 62.0 $\pm$ 1.4 kg, respectively. The amputation side for one of the amputees is right, and the others are left. All experiments were approved and performed under the supervision of the SUSTech (Southern University of Science and Technology) Medical Ethics Committee (approval number: 20210009, date:2021/3/2).

III. Results

A. Network Training Results

The classification accuracy on the validation set of each epoch of the designed neural network are shown in Fig. 6. The results indicates that the average classification accuracy on the validation set in the last 30 epochs (70 - 100) reaches 94.70%. In addition, after 20 epochs, the loss of both classifiers and the discrepancy between classifiers decreases to less than 0.01. In the last thirty epochs, the average loss and discrepancy of the two classifiers are 0.0029 and 0.0017, respectively.

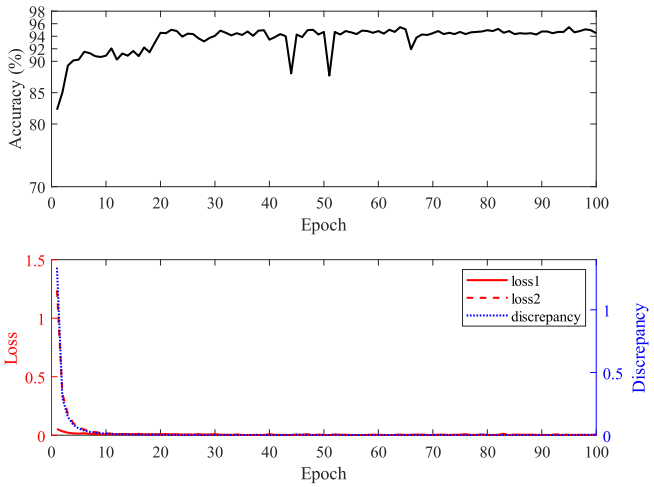


Fig. 6. Accuracy, loss and discrepancy between the two classifiers of each epoch of the proposed adaptation network. Classification accuracy is tested on the manually labeled validation set. Loss1 and loss2 are the loss of classifier C_1 and C_2 respectively. Discrepancy is calculated after the third step in each epoch.

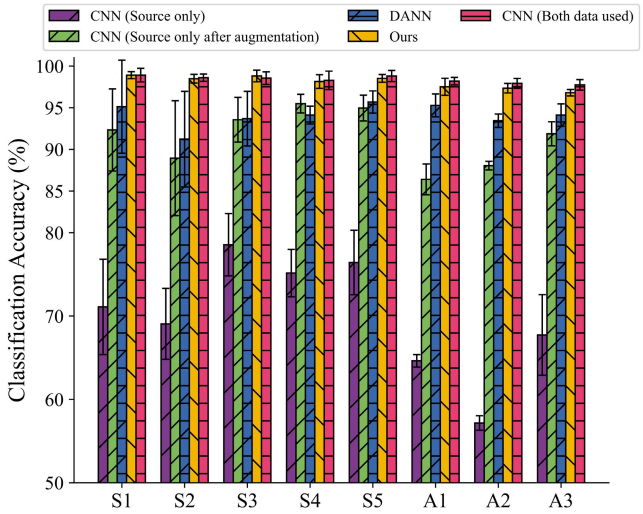


Fig. 7. Indoor classification accuracy for each subject.S1-S5 represent five able-bodied subjects. A1-A3 represent three amputees.

Indoor Experiment		Predicted Modes (%)				
		LG	SA	SD	RA	RD
Actual Modes (%)	LG	97.07	2.23	0.33	0.13	0.23
	SA	1.10	98.90	0.00	0.00	0.00
	SD	1.74	0.05	98.21	0.00	0.00
	RA	1.40	0.18	1.04	97.28	0.10
	RD	1.11	0.00	0.00	0.92	97.97

Fig. 8. Indoor classification confusion matrix for all subjects.

B. Indoor Experiment Results

The indoor environmental data is used to validate the classification performance of the proposed sim-to-real transfer approach secondly. The classification of each sample takes about 6ms. Since the main objective of this study is to accurately classify real-world terrains, and the geometries of different kinds of simulated terrains vary greatly, only the performance on real-world data is evaluated. As shown in Fig. 7, after adaptation, the average classification accuracy for all subjects on the five environments in the indoor experiment is 98.06% ± 0.71% (98.57% and 97.22% for able-bodied subjects and amputees). A t-test and a one-way ANOVA with post hoc test at a significance level of $\alpha = 0.05$ are used to compare the difference in results between different methods. Besides, a P value is used to denote the probability that the null hypothesis is true. The results show that, the adaptation network performed significantly better than CNN trained with simulated data only (69.98% ± 6.54% and 91.45% ± 3.11% for before and after data augmentation of simulation dataset, respectively) ($P < 0.001$) and DANN (94.43% ± 1.60%) ($P < 0.01$). In addition, the lower standard deviation of adaptation network presented its better generalization ability to different subjects than CNN trained with simulated data only and DANN.

Moreover, the average accuracy of CNN trained with labeled real-world data and simulated data using supervised learning is also calculated and considered as an upper bound. This CNN achieves a 0.31% higher accuracy than the adaptation network, but their difference is not significant (P = 0.33). Confusion matrix is used to further evaluate the classification performance (see Fig. 8). The accuracy is relatively high for stairs (98.90% and 98.21% for stair ascent and descent).

In addition, to validate whether the camera placed on different leg or different gender of a subject will influence the experimental result, we invited six more able-bodied subjects to perform indoor walking to collect environmental data and conducted an offline classification experiment. The results indicate that, there is no significant difference between different genders of the subject ($P = 0.54$) and between different legs the camera is placed on ($P = 0.75$). More details of the experimental setup and experimental results are presented in the supplementary document attached with the paper.

C. Outdoor Experiment Results

To verify the generalization capability of this method in outdoor environment, the performance of outdoor classification experiment is evaluated. The classification of each sample takes about 6ms. As shown in Fig. 9, the average accuracy in outdoor environments is 95.91% ± 1.09% (95.25% for able-bodied subjects and 97.01% for amputees). Since simulating outdoor environment present more difficulties, the results of adaptation network present a more significant difference between those of CNN trained with simulated data only (46.41% and 23.92% higher than before and after simulated data augmentation respectively) ($P < 1 \times 10^{-4}$). The proposed method also outperforms DANN (87.75% ± 3.22%) ($P = 2.28 \times 10^{-4}$). In addition, the adaptation network still presents a low standard deviation (1.09%), demonstrating its capability of generalize to different subjects in outdoor environment. The average accuracy of CNN trained with supervised learning using all data (97.05% ± 1.00%) is slightly higher

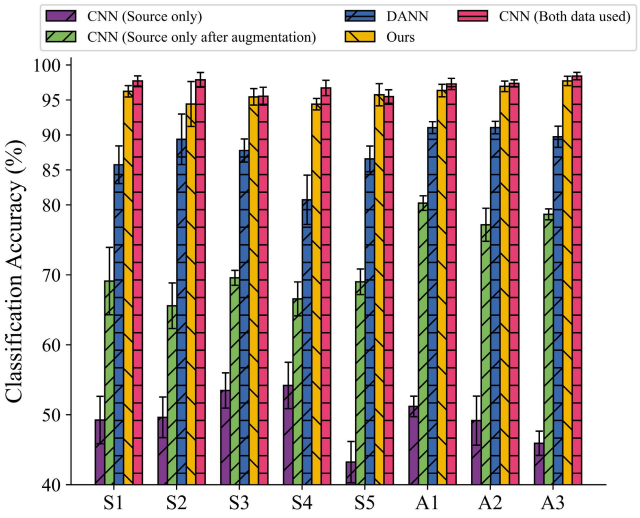
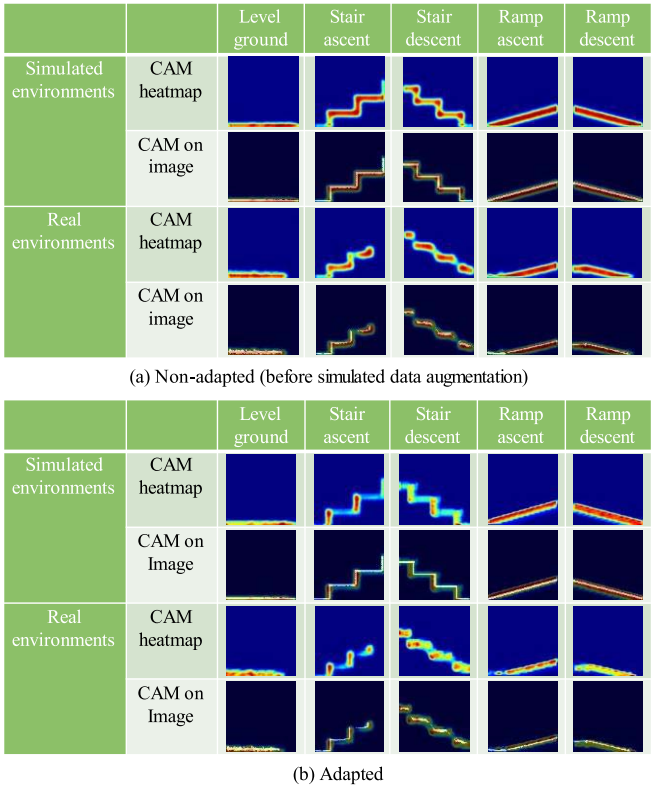


Fig. 9. Outdoor classification accuracy for each subject. S1-S5 represent five able-bodied subjects. A1-A3 represent three amputees.

Outdoor Experiment		Predicted Modes (%)				
		LG	SA	SD	RA	RD
Actual Modes (%)	LG	91.01	0.48	0.62	2.00	5.89
	SA	1.05	97.50	0.01	1.43	0.00
	SD	1.26	0.01	97.81	0.00	0.92
	RA	3.67	0.29	0.00	96.01	0.03
	RD	1.35	0.00	0.00	0.16	98.39

Fig. 10. Outdoor classification confusion matrix for all subjects.



(a) Non-adapted (before simulated data augmentation)

(b) Adapted

Fig. 11. (Best viewed in color) CAM visualization of the hidden features extracted from different kinds of simulated and real-world environmental images before and after simulated data augmentation and domain adaptation. The color of the regions in the figure from blue-green-yellow-red represents the contribution to the classification task from small to large. (The red region represents the extracted features with the greatest contribution to the classification task.)

than our adaptation network, but their difference remains insignificant ($P = 0.06$).

Moreover, the confusion matrix for outdoor experiment (see Fig. 10) is evaluated, indicating increased misclassifications in outdoor experiments. Compared with other terrains with relatively distinctive features, classification of level ground is more susceptible to outdoor interfering factors such as sunlight and complex road conditions. This may lead to the greater variation in classification accuracy of level ground between indoor (97.07%) and outdoor (91.01%).

D. Visualization Results

The CAM visualization results of the essential features of different environments extracted by the adapted and non-adapted feature extractor are shown in Fig. 11. The red region in the figure indicates the region that contributes most to the classification tasks, explaining the classification mechanism. As shown in Fig. 11(b), the adapted extractor is able to extract essential features of the terrain from different images and distinguish them from the rest of the image, which is also positively correlated with classification accuracy. For stairs, the adapted network mainly extracts the part with sudden change in height for classification, i.e., the vertical surface of stairs. In addition, for stair descent with discontinuous parts in real world, the corner points of the stair are mainly extracted. For other terrains, the classification task is also based on the changes in height. Terrain with continuous large changes in height is classified as ramp, and terrain with constant or with continuous small changes in height is classified as level ground. However, as shown in Fig. 11(a), the non-adapted extractor presents difficulties in distinguishing the most important features for classification with other parts in environmental images, which may be related to its inferior performance in classifying more unstructured real-world environments.

TABLE I
COMPARISON WITH PREVIOUS RESEARCHES

Parameters	[14]	[15]	[19]	Ours
Method	SVM	CNN	Bayesian neural networks	Unsupervised sim-to-real adaptation
Using Labeled Real-world Data for Training	Yes	Yes	Yes	No
Classification Accuracy	94.1% (indoor)	94.85% (indoor and outdoor)	95.15% (indoor and outdoor)	98.06% (indoor) 95.91% (outdoor)
Environment Modes	3	3	6	5
Subject Number and Type	Four able-bodied subjects	One able-bodied subject	Seven able-bodied subjects and one amputee	Six able-bodied subjects and three amputees
Real-world Dataset Size	402,403 depth and confidence images	34254 RGB images	327,000 RGB images	7500 binary images

To visualize and validate the effect of the proposed sim-to-real adaptation approach on feature distributions from the source domain and the target domain, the features extracted by the last hidden layer of the adapted and non-adapted feature extractor are projected onto the 2D plane by t-SNE. As shown in Fig. 12(a), before adaptation, there is almost no overlap between the source features (dark-colored points) and the target (light-colored points) and the features of the same terrain in the target domain are dispersed, which may be related to the misclassification. Although the augmentation of simulation dataset helps the network to better cluster the target features (see Fig. 12(b)), the features in the source and target domain are still not well aligned. However, as shown in Fig. 12(c), domain adaptation effectively reduces the distance between the source domain and target domain and presents a good clustering effect on the target data. In addition, to quantify the effect of domain adaptation on feature alignment, the total distance between the center of the source features and the target features (indicated as d_1) before and after adaptation is calculated. The results show that, after adaptation, d_1 decreases from 2.68 (in Fig. 12(a)) to 0.54 (Fig. 12(c)), indicating the success of feature alignment and a high real-world classification accuracy.

IV. Discussion

In this study, an unsupervised sim-to-real adaptation method is proposed for precise real-world environmental classification, thus providing environmental context for human motion intent prediction and achieving sim-to-real transfer. Compared with existing environmental recognition studies based on supervised learning (see Table I), the major advantage of the proposed approach is maintaining a high real-world classification accuracy while avoiding the burden of data annotation. The experimental environments and datasets in Table I are different. These results cannot be used to quantitatively compare different methods, but can be used as a reference to qualitatively show the performance of our method.

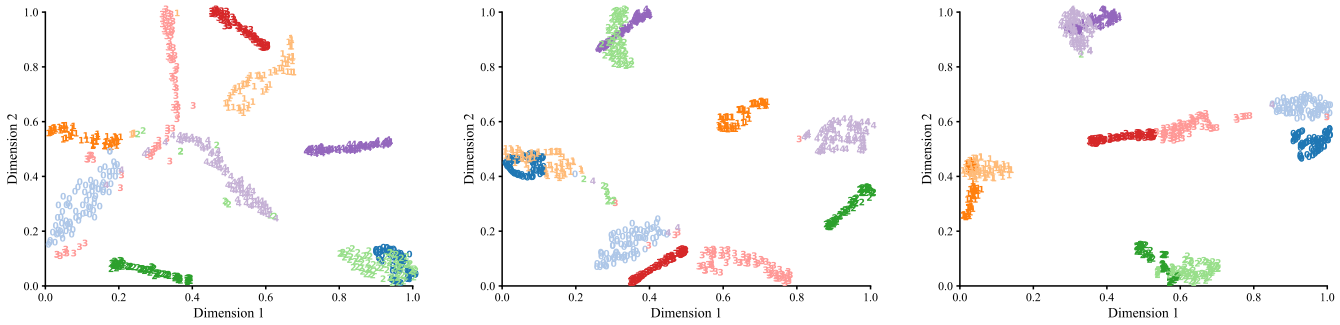
The results of indoor and outdoor classification experiments indicate that the trained adaptation network can accurately identify the environment in both indoor and outdoor environments (98.06% and 95.91%). This is about 6% higher than DANN and similar to supervised learning using both types of data ($P = 0.33$ and $P = 0.06$ in indoor and outdoor experiments), which is regarded as an upper bound of the proposed approach. Because the main goal of this study is to achieve sim-to-real adaptation, the similar accuracy and the visualization performance are sufficient to demonstrate that the trained network has successfully aligned source and target features and mitigated the sim-to-real gap. Besides, our approach achieves outstanding performance on all subjects (accuracy higher than 94% for each subject) with a low standard deviation (0.71% and 1.09% in indoor and outdoor experiments), demonstrating the generalization capability to different subjects. The time to convert the point cloud to binary image and the time to classify the image for each frame were also calculated by a computer with the same configuration as in Section II.C, which took about 20ms and 6ms respectively. Therefore, the total processing time for each frame was about 26 ms, which was shorter than the acquisition time of the depth camera (> 30 ms) and was suitable for real-time implementation.

In addition, the proposed method avoids the time-consuming and laborious data annotation compared with previous supervised learning-based studies. In previous researches, to improve the accuracy and the generalization capability of the network, researchers usually need to collect and label tens to hundreds of thousands of images for training. Zhong *et al.* [19] and Massalin *et al.* [14] annotated about 327,000 RGB images and about 400,000 depth and confidence images in their studies, respectively. The largest dataset of environmental recognition for prostheses and exoskeletons is the ExoNet proposed by Laschowski *et al.* [23]. To develop a large-scale database for recognizing human walking environments, they collected approximately 5.6 million images, and manually labeled about 923,000 images. The proposed method circumvents this burden by using simulated data and unsupervised domain adaptation.

This paper also investigates the classification performance of the adaptation network trained with less target data. The results show that, when training with only 10% of target data (750 unlabeled images), the network still obtains an accuracy of 96.34% and 91.09% in indoor and outdoor environments respectively, which is still higher than DANN. This indicates the proposed method requires less real-world data and has the potential to be more easily generalized to other environments such as transition states. The transition states between different terrains are more challenging to classify than the steady-state terrains mainly focused on in this paper [39]. When facing such situations, supervised learning methods need to re-collect and re-annotate environmental data, while the proposed method is expected to collect less real-world data and avoid the labeling task by generating corresponding simulated environmental data.

Moreover, many machine learning-based environmental recognition methods have achieved excellent performance, but few have explained their basis for classification. In this study, the classification mechanism is explicitly explained by visualizing the essential features of different terrains extracted by the trained feature extractor. The results indicate that the network's classification of different terrains is mainly depends on the identified height change of the terrain. This is also intuitively consistent with the logic of human classification of these five terrains. Moreover, the visualization results can help improve the understanding of CNN, and are expected to guide the generation of more realistic simulated data.

Although the proposed method successfully realizes accurate real-world environmental classification and mitigates the sim-to-real-gap, there are still some limitations. First, this method is only analyzed offline and the participating amputees perform the walking experiment with passive prostheses. When walking with a powered prosthesis, the amputee may perform a different gait, which may present new challenges. Therefore, further online evaluation of powered prostheses is required. Second, the generated simulated



(a) Non-adapted (b) Non-adapted (after data augmentation) (c) Adapted

Fig. 12. (Best viewed in color) T-SNE visualization of the hidden features for simulated and real-world data randomly selected from source and target dataset before and after adaptation. Deep color and light color points of the same color family represent the source and the target hidden features from the same class. The different numbers denote different classes of the hidden features.

point cloud considers real-world unstructured terrain and camera perspective. However, amputees will inevitably encounter obstacles, pedestrians, and other unexpected reality factors affecting the classification during real-world walking. The geometries of these factors are difficult to define. In addition, the transition states between different terrains are not explicitly classified in this paper. Therefore, future work may take the classification of transition states into account and utilize a more photo-realistic simulator such as Grand Theft Auto V (GTA-V) [40] to generate different terrains with more realistic factors for more detailed environment classification and further reduction of real-world data demand. Finally, environmental recognition can be helpful for motion intent prediction, but it cannot replace the prediction. Therefore, the proposed method needs to be applied jointly with intent prediction methods [41] to enhance the multi-environmental real-time control of the prostheses.

V. Conclusion

The sim-to-real gap forms the barrier to the utilization of simulated data to overcome the labeled real-world data requirement for supervised learning-based environmental recognition. In this study, an unsupervised sim-to-real adaptation approach is developed to address this problem. The proposed approach only utilized simulated data and unlabeled real-world environmental data to train a network to accurately classify real-world terrains and mitigate the sim-to-real gap. This method is evaluated by inviting subjects to perform indoor and outdoor walking experiments to capture environmental data. According to experimental results, the proposed approach successfully realizes accurate real-world environmental classification (average accuracy: 98.06% $\pm$ 0.71% and 95.91% $\pm$ 1.09% in indoor and outdoor environments). The results are also close to the upper bound, which is the result of CNN trained with labeled real-world data and simulated data (98.37% and 97.05% in indoor and outdoor environment), achieving the objective of this paper. Moreover, by visualizing the features of different terrains extracted by the network, visual interpretation of classification mechanism of the network is provided, which is based on the height change of the terrain. This study demonstrates that the proposed method can effectively mitigate the sim-to-real gap and provide environmental context for human motion intent prediction during the control of powered prostheses and human-robot interaction.

References

[1] K. Ziegler-Graham, E. J. MacKenzie, P. L. Ephraim, T. G. Travison, and R. Brookmeyer, "Estimating the prevalence of limb loss in the United States: 2005 to 2050," *Archi. Phys. Med. Rehabil.*, vol. 89, no. 3, pp. 422–429, Mar. 2008.

[2] A. Seker, A. Kara, S. Camur, M. Malkoc, M. M. Sonmez, and M. Mahirogullari, "Comparison of mortality rates and functional results after transtibial and transfemoral amputations due to diabetes in elderly patients—A retrospective study," *Int. J. Surg.*, vol. 33, pp. 78–82, Sep. 2016.

[3] B. E. Lawson, J. Mitchell, D. Truex, A. Shultz, E. Ledoux, and M. Goldfarb, "A robotic leg prosthesis: Design control and implementation," *IEEE Robot. Autom. Mag.*, vol. 21, no. 4, pp. 70–81, Dec. 2014.

[4] D. Quintero, D. J. Villarreal, D. J. Lambert, S. Kapp, and R. D. Gregg, "Continuous-phase control of a powered Knee–Ankle prosthesis: Amputee experiments across speeds and inclines," *IEEE Trans. Robot.*, vol. 34, no. 3, pp. 686–701, Jun. 2018.

[5] S. K. Au, J. Weber, and H. Herr, "Powered ankle-foot prosthesis improves walking metabolic economy," *IEEE Trans. Robot.*, vol. 25, no. 1, pp. 51–66, Feb. 2009.

[6] 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.

[7] K. Zhang, J. Chen, J. Wang, Y. Leng, C. W. de Silva, and C. Fu, "Gaussian-guided feature alignment for unsupervised cross-subject adaptation," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108332.

[8] F. Gao, G. Liu, F. Liang, and W.-H. Liao, "IMU-based locomotion mode identification for transtibial prostheses, orthoses, and exoskeletons," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1334–1343, Jun. 2020.

[9] J. S. Matthis, J. L. Yates, and M. M. Hayhoe, "Gaze and the control of foot placement when walking in natural Terrain," *Current Biol.*, vol. 28, no. 8, pp. 1224–1233, Apr. 2018.

[10] Y. Hu, Z. Li, G. Li, P. Yuan, and C. Yang, "Development of sensory-motor fusion-based manipulation and grasping control for a robotic hand-eye system," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 47, no. 7, pp. 1169–1180, Jul. 2017.

[11] K. Zhang *et al.*, "Foot placement prediction for assistive walking by fusing sequential 3D gaze and environmental context," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2509–2516, Apr. 2021.

[12] Y. Qian *et al.*, "Predictive locomotion mode recognition and accurate gait phase estimation for hip exoskeleton on various terrains," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 6439–6446, Jul. 2022, doi: 10.1109/LRA.2022.3173426.

[13] B. Laschowski, W. McNally, A. Wong, and J. McPhee, "Comparative analysis of environment recognition systems for control of lower-limb exoskeletons and prostheses," in *Proc. 8th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Nov. 2020, pp. 105–110.

[14] 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.

[15] B. Laschowski, W. McNally, A. Wong, and J. McPhee, "Preliminary design of an environment recognition system for controlling robotic lower-limb prostheses and exoskeletons," in *Proc. IEEE 16th Int. Conf. Rehabil. Robot. (ICORR)*, Jun. 2019, pp. 868–873.

[16] 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.

[17] G. Tian, J. Chen, X. Zeng, and Y. Liu, "Pruning by training: A novel deep neural network compression framework for image processing," *IEEE Signal Process. Lett.*, vol. 28, pp. 344–348, 2021.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[19] B. Zhong, R. L. D. Silva, M. Li, H. Huang, and E. Lobaton, "Environmental context prediction for lower limb prostheses with uncertainty quantification," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 458–470, Apr. 2021.

[20] 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.

[21] K. Zhang *et al.*, "A subvision system for enhancing the environmental adaptability of the powered transfemoral prosthesis," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3285–3297, Jun. 2021.

[22] K. Bousmalis *et al.*, "Using simulation and domain adaptation to improve efficiency of deep robotic grasping," in *Proc. IEEE Int. Conf. Robot. Autom.*, Brisbane, QLD, Australia, May 2018, pp. 4243–4250.

[23] B. Laschowski, W. McNally, A. Wong, and J. McPhee, "ExoNet database: Wearable camera images of human locomotion environments," *Frontiers Robot. AI*, vol. 7, p. 188, Dec. 2020.

[24] B. Laschowski, W. McNally, A. Wong, and J. McPhee, "Environment classification for robotic leg prostheses and exoskeletons using deep convolutional neural networks," *Frontiers Neurorobot.*, vol. 15, Feb. 2022, Art. no. 730965.

[25] N. Sünderhauf *et al.*, "The limits and potentials of deep learning for robotics," *Int. J. Robot. Res.*, vol. 37, nos. 4–5, pp. 405–420, 2018.

[26] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2017, pp. 23–30.

[27] F. Zhu, L. Zhu, and Y. Yang, "Sim-real joint reinforcement transfer for 3D indoor navigation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11380–11389.

[28] Z. Liu, Z. Meng, N. Gao, and Z. Zhang, "Calibration of the relative orientation between multiple depth cameras based on a three-dimensional target," *Sensors*, vol. 19, no. 13, p. 3008, Jul. 2019.

[29] S. Lee and H. Shim, "Skewed stereo time-of-flight camera for translucent object imaging," *Image Vis. Comput.*, vol. 43, pp. 27–38, Nov. 2015.

[30] 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.

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

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

[33] X. Wang, L. Li, W. Ye, M. Long, and J. Wang, "Transferable attention for domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 5345–5352.

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

[35] 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.

[36] H. Wang *et al.*, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 24–25.

[37] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. 29th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2921–2929.

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

[39] B.-Y. Su, J. Wang, S.-Q. Liu, M. Sheng, J. Jiang, and K. Xiang, "A CNN-based method for intent recognition using inertial measurement units and intelligent lower limb prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 1032–1042, May 2019.

[40] W. Bichen, W. Alvin, X. Yue, and K. Keutzer, "SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2018, pp. 1887–1893.

[41] Y.-X. Liu, R. Wang, and E. M. Gutierrez-Farewik, "A muscle synergy-inspired method of detecting human movement intentions based on wearable sensor fusion," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1089–1098, 2021.