2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Alleviating Feature Confusion in Cross-Subject Human Activity Recognition via Adversarial Domain Adaptation Strategy

Yalan Ye[1,*], Qiang Zhou[1], Tongjie Pan[1], Ziwei Huang[1] and Zhengyi Wan[1]

Abstract— **Sensor-based Human Activity Recognition (HAR) plays an important role in health care. However, great individual differences limit its application scenarios and affect its performance. Although general domain adaptation methods can alleviate individual differences to a certain extent, the performance of these methods is still not satisfactory, since the feature confusion caused by individual differences tends to be underestimated. In this paper, for the first time, we analyze the feature confusion problem in cross-subject HAR and summarize it into two aspects: Confusion at Decision Boundaries (CDB) and Confusion at Overlapping (COL). The CDB represents the misclassification caused by the feature located near the decision boundary, while the COL represents the misclassification caused by the feature aliasing of different classes. In order to alleviate CDB and COL to improve the stability of trained model when processing the data from new subjects, we propose a novel Adversarial Cross-Subject (ACS) method. Specifically, we design a parallel network that can extract features from both image space and time series simultaneously. Then we train two classifiers adversarially, and consider both features and decision boundaries to optimize the distribution to alleviate CDB. In addition, we introduce Minimum Class Confusion loss to reduce the confusion between classes to alleviate COL. The experiment results on USC-HAD dataset show that our method outperforms other generally used cross-subject methods.**

I. INTRODUCTION

Compared to visual-based Human Activity Recognition (HAR), sensor-based HAR has been widely used in health monitoring [1], elderly care [2], smart home [3], etc. due to its greater convenience and better privacy protection. However, the recognition accuracy suffers a major setback because of the physiological and behavioral differences between subjects, like age, Body Mass Index (BMI), behavior habit, etc. Although this issue can be solved by training a large number of users' data to obtain a more generalized model, data collection and labeling can be tedious, time-consuming, and even difficult to accomplish in some cases (e.g. collecting vast data from patients with rare diseases). Therefore, it is a meaningful and challenging task to solve cross-subject problems in HAR.

To our knowledge, there are few studies on cross-subject HAR, and existing solutions are extremely based on domain adaptation strategy. Domain adaptation aims to transfer knowledge learned from source domain to the target [4], so the general domain adaptation methods are often used to solve cross-subject problems. Some domain adaptation methods are based on metric learning, such as Maximum Mean Discrepancy (MMD). The experiment results of Ding et al. [5] show the effectiveness of MMD in cross-subject HAR. Another domain adaptation methods are based on adversarial idea, which is inspired by Generative Adversarial Networks (GAN) [6]. Based on the structure of traditional GAN, Soleimani et al. [7] proposed Subject Adaptor GAN (SA-GAN) to solve cross-subject HAR problem. Another adversarial domain adaptation method, Domain-Adversarial Neural Network (DANN) [8], is also used in [5]. However, the performance of these methods is not satisfactory when there are large differences between individuals. Because one of the basic assumptions for general domain adaptation methods is that features of the same category are closer between the source and the target [9]. While in cross-subject HAR scenario, the large differences between individuals can make feature confusion worse, which may not satisfied the basic assumption.

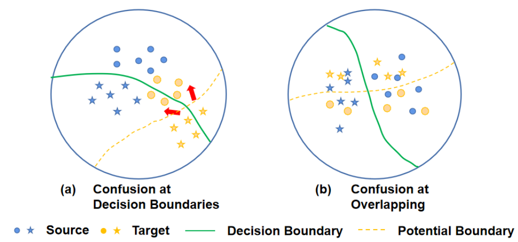


Fig. 1. Two aspects of the feature confusion problem. (a): Features near the decision boundary may be misclassified during alignment. Red arrows indicate the direction of feature movement when aligning. (b): The overlapping features of different classes in source and target domain lead to confusion, while they are distinguishable in their respective domains.

The feature confusion problem has two aspects, which is shown in Fig.1. One is that the features of the target individuals are distributed near the decision boundaries, which may cause misalignment. This can be defined as Confusion at Decision Boundaries (CDB) (shown in Fig.1 (a)). The other is that the features of different categories in source and target domain are mixed together, leading to misclassification. This can be defined as Confusion at Overlapping (COL) (shown in Fig.1 (b)).

In this paper, we propose a novel Adversarial Cross-Subject (ACS) method to alleviate the feature confusion. We specially design a parallel network for the cross-subject HAR task, which can simultaneously extract features from both image space and time series. As an adversarial method, we

*This work was supported in part by a grant from the National Natural Science Foundation of China (No.61976047), in part by grants from Science & Technology Department of Sichuan Province of China (No. 2020YFG0087, 2020YFG0326 and 2021YFG0331), in part by a grant from the Fundamental Research Funds for the Central Universities (No. ZYGX2021YGLH016), and in part by a grant from the Open Innovation Fund of 55 Institute of China North Industries Group.

[1]Yalan Ye, Qiang Zhou, Tongjie Pan, Ziwei Huang and Zhengyi Wan are with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China.

*Corresponding author: Yalan Ye yalanye@uestc.edu.cn

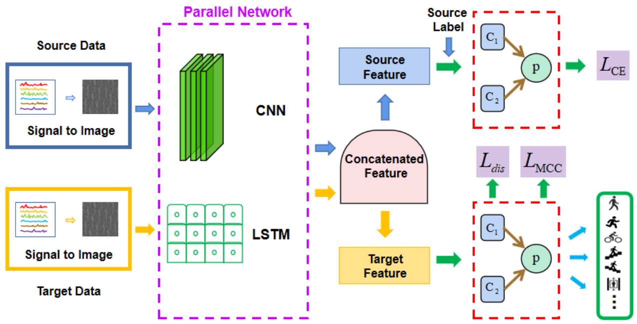


Fig. 2. The framework of our ACS method. The signal data is transformed into images as the input of our method. The purple dotted box is the parallel network specially designed for the cross-subject HAR task. Two classifiers and MCC loss are used to optimize the distribution of extracted features to alleviate CDB and COL.

also incorporate the idea of adversarial training to optimize the distribution. We adversarially train two classifiers to reduce the number of confusable features near the decision boundaries, which can alleviate the effect of CDB. And to alleviate COL, we introduce Minimum Class Confusion (MCC) loss [10] into our method. By minimizing the degree of confusion between classes of the confusion matrix, the effect of COL can be reduced.

Our main contributions in this paper are summarized as follows:

- To reveal the main difficulties in cross-subject HAR, we analyze the problem of feature confusion for the first time, which includes Confusion at Decision Boundaries (CDB) and Confusion at Overlapping (COL).
- We propose a novel ACS method, which can optimize the distribution of signal features to alleviate CDB and COL and improve the stability of the model in cross-subject HAR.
- To alleviate CDB and COL, we design a parallel network for the cross-subject HAR task specially, which can extract features from both image space and time series simultaneously to obtain more stable features.
- We demonstrate the effectiveness of our method on USC-HAD dataset, and it has higher average accuracy on target subjects than other generally used domain adaptation methods.

II. The Framework of Our Proposed Method

In this section, we present the detail of our proposed method. Firstly, we introduce the parallel network we designed for the cross-subject HAR task. Then we describe the training process of our proposed method in detail and explain how we alleviate the effects of CDB and COL. Fig.2 shows the overall framework of our method.

A. Proposed Parallel Network

Feature extraction network plays a crucial role in HAR tasks. Single dimensional features are more likely to be influenced by individual difference. Therefore, we design a parallel network to extract features from both image space and time series simultaneously. The input of the parallel network is a single channel image which is formed by directly stacking signal channels. It is similar to the Activity

TABLE I

The Structure and Parameters of Parallel Network

Network	Structure and Parameters						
CNN	Block1	Conv2d	in_c	out_c	k_size	s	p
			1	128	3	1	1
		MaxPool2	/	/	2x1	1	0
	Block2	Conv2d	in_c	out_c	k_size	s	p
			128	256	3	1	1
		MaxPool2d	/	/	2x1	1	0
	Block3	Conv2d	in_c	out_c	k_size	s	p
			256	64	3	1	0
		MaxPool2d	/	/	2x1	2x1	0
LSTM	hidden_size		num_layers				
	64		3				

Image proposed in [11]. Considering the advantages of Convolutional Neural Network (CNN) in image processing, we use it to extract the features of image space, so as to obtain the potential connections between various signal channels. In addition, we use Long Short-Term Memory (LSTM) network to extract features from time series due to the time continuity of human activities. Table I shows the specific structure and parameters of the parallel network. For each block of CNN, a batch normal layer and a ReLU layer are added between the convolution layer and the pooling layer. After passing through our specially designed parallel network, the features of image space and time series are extracted.

B. Training Steps

Based on Maximum Classifier Discrepancy (MCD) strategy [12], we divide the whole training process into three interrelated steps.

1) Step one: Step one is to use the source domain data to train the parallel network E and classifier C_1 and C_2. The purpose is to enable the model to classify the source domain samples correctly to obtain task-specific decision boundaries. If we use (x, y) to denote sample data, (X, Y) to denote actual data distribution, and subscripts s and t to distinguish source domain and target, then the objective of this step is as follows:

$$\mathcal{L}_{CE} = -\mathbb{E}_{(x_s,y_s)\sim(X_s,Y_s)} \sum_{n=1}^{2}\sum_{m=1}^{M} \mathbb{I}_{[m=y_s]} \log p_n(y|x_s), \quad (1)$$

$$\mathcal{L}_{step1} = \mathcal{L}_{CE}, \quad (2)$$

where $\mathcal{L}_{CE}$ denotes cross entropy loss, M denotes the number of classes, and $p_n(y|x_s)$ denotes the prediction probability of classifier C_n.

2) Step two: Step two is to maximize the discrepancy between two classifiers with E fixed. The discrepancy loss $\mathcal{L}_{dis}$ is defined as follows:

$$\mathcal{L}_{dis} = \frac{1}{M}\sum_{m=1}^{M} |p_{1m}(y|x_t) - p_{2m}(y|x_t)|, \quad (3)$$

where $p_{1m}(y|x_t)$ and $p_{2m}(y|x_t)$ denote probability output of C_1 and C_2 for class m on x_t respectively. However, it should be noted that while enlarging the discrepancy between C_1 and C_2, their classification effect on x_s should be ensured. Therefore, we add $\mathcal{L}_{CE}$ in this step. Thus, the final target loss function of step two is as follows:

$$\mathcal{L}_{step2} = \mathcal{L}_{CE} - \mathcal{L}_{dis}. \quad (4)$$

3) Step three: Step three is to train E to obtain features based on new decision boundaries with C_1 and C_2 fixed. This step uses x_t only, whose purpose is to align the extracted target features to the source domain. We do this by minimizing the discrepancy loss $\mathcal{L}_{dis}$, which constitutes a minimax game with the step two.

In order to reduce the effect of COL at the same time, we introduce MCC loss. MCC loss is a universal domain adaptation loss function, which can be used to solve the problem of class confusion. It is defined as follows:

$$\mathcal{L}_{MCC}(\hat{Y}_t) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j\neq i}^{M}|\tilde{H}_{ij}|, \tag{5}$$

where $\hat{Y}_t = f_\theta(X_t) \in \mathbb{R}^{B\times M}$ and $\tilde{H}_{ij}$ denotes the confusion degree between class i and class j. Here we use f_θ to denote our model and B to denote the size of each batch. $\tilde{H}_{ij}$ can be calculated by $\hat{Y}_t$ and $\hat{Y}_t{}^{\top}$ (Specific definition can refer to [10]). Therefore, the objective of this step is as follows:

$$\mathcal{L}_{step3} = \beta(\mathcal{L}_{dis} + \alpha\sum_{n=1}^{2}\mathcal{L}_{MCC}(\hat{Y}_{nt})), \tag{6}$$

where β is the weight of step three and α is the weight of $\mathcal{L}_{MCC}$. They are both hyper-parameters in our experiments. Subscripts n corresponds to classifier C_n.

III. Experiments and Results

We mainly do experiments on USC-HAD [13], which is a public dataset using 3-axis acceleration and 3-axis gyroscope signals for HAR tasks. This dataset contains data from 14 subjects, each of whom performed 12 different activities. Table II lists the relevant information for each subject in the USC-HAD dataset.

TABLE II

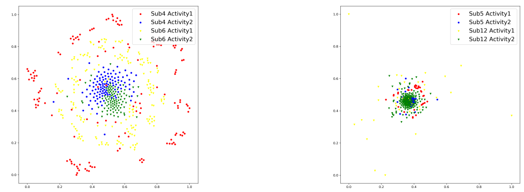
Subject Information in USC-HAD Dataset

Subject	Age	Height(cm)	Weight(kg)	BMI(kg/m^2)
Sub.1	27	164	43	16.0
Sub.2	26	185	75	21.9
Sub.3	31	169	68	23.8
Sub.4	23	168	52	18.4
Sub.5	35	170	63	21.8
Sub.6	27	164	50	18.6
Sub.7	32	160	75	29.3
Sub.8	22	180	76	23.5
Sub.9	30	171	60	20.5
Sub.10	28	170	75	26.0
Sub.11	34	165	48	17.6
Sub.12	36	170	80	27.7
Sub.13	21	178	71	22.4
Sub.14	49	166	68	24.7

[1] The normal BMI range for adults is 18.5 to 23.9.

We mainly studies the feature confusion caused by individual differences in this paper. Therefore, when setting up sub-experiments, our principle is to make the coverage of individual differences among subjects as wide as possible. Finally, Subjects 4, 5, 6, 7, 12, and 14 are selected for subsequent cross-subject experiments, which included both individuals with similar physical signs and individuals with large differences. Taking the groups of Sub.4, Sub.6 and Sub.5, Sub.12 as examples, Fig.3(a) and Fig.3(b) show the performance of individual differences at the feature level.

In the data preprocessing stage, we reduced the sampling rate to 20Hz and selected a 5-second sliding window with 3-second overlap. All signal channels are simply stacked together to form a simplified single channel activity image. In this way, the original signal slice is converted into a 100×6 image. We combined "elevator up" and "elevator down" into one activity in our experiments. This kind of processing is reasonable in practical scene, because in most cases we do not need to accurately understand the running direction of the elevator, and the information contained in a 5-second time slice is not enough to accurately distinguish them. Therefore, the experiments enumerated in this paper are all 11-classification tasks.



(a) Mild confusion (b) Severe confusion

Fig. 3. The performance of individual differences at the feature level. (a): Mild feature confusion between similar individuals. (b): Severe feature confusion when individual differences are large.

To verify the effectiveness of the parallel network, we compared the activity recognition accuracy on target subjects of using only single-type networks without any cross-subject methods, as shown in Table III. The results show that the parallel network we designed has higher average accuracy on target subjects than single-type networks, which proves the potential ability of our parallel network in alleviating CDB and COL.

TABLE III

Average Accuracy of Different Networks on Target Subjects

Network	Method	Average Accuracy
CNN only	Only Source	0.6247
LSTM only	Only Source	0.5321
Parallel Network(Proposed)	Only Source	0.6484

The cross-subject experiment results are shown in Table IV. The ratio of training set to test set is 8:2. We compared the effect of our method with several other widely used domain adaptation methods, such as MMD, Conditional Domain Adversarial Network (CDAN) [14], DANN and SA-GAN. In order to obtain more reliable results, we recorded five optimal values for each group of experiments and calculated their mean and standard deviation. The results show that out of all thirty sub-experiments, our method achieved the best results in 25 groups and achieved the second best results in 2 groups. In terms of average accuracy, our method improves the effect by 11.69%, which is much

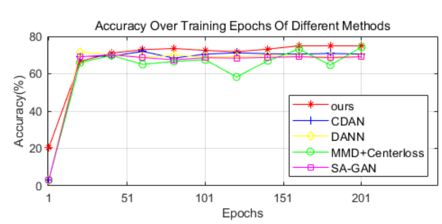


Fig. 4. Classification accuracy over training epochs of different methods. The curves of other methods fluctuate greatly or the optimal value appears at an earlier time.

TABLE IV

RESULTS OF SSDA EXPERIMENTS ON USC-HAD DATASET

Source	Target	Only Source (Baseline)	CDAN[14]	DANN[8]	MMD+ Centerloss[5]	SA-GAN[7]	**ACS (ours)**
Sub.4	Sub.5	0.8174±0.0210	0.8137±0.0084	0.8575±0.0109	0.8072±0.0237	0.8065±0.0125	**0.8867±0.0025**
	Sub.6	0.7991±0.0360	0.8138±0.0147	0.8012±0.0544	0.7815±0.0327	0.7683±0.0312	**0.8683±0.0028**
	Sub.7	0.7519±0.0182	0.7822±0.0034	0.7890±0.0022	0.7769±0.0141	0.7757±0.0146	**0.8920±0.0008**
	Sub.12	0.5922±0.0094	0.6529±0.0077	0.5856±0.0060	0.4917±0.0136	0.5817±0.0092	**0.7964±0.0067**
	Sub.14	0.4925±0.0101	0.5797±0.0040	0.5642±0.0273	0.4595±0.0236	0.5501±0.0225	**0.6650±0.0072**
Sub.5	Sub.4	0.7791±0.0446	0.8627±0.0114	0.8801±0.0023	0.8041±0.0200	0.8094±0.0219	**0.9255±0.0063**
	Sub.6	0.8138±0.0265	0.8675±0.0071	0.8651±0.0027	0.6909±0.0117	0.7199±0.0094	**0.8871±0.0053**
	Sub.7	0.8129±0.0181	0.8774±0.0033	0.8788±0.0034	0.8540±0.0299	0.8455±0.0176	**0.9073±0.0073**
	Sub.12	0.5723±0.0267	0.5414±0.0016	0.6312±0.0023	0.5320±0.0123	0.5615±0.0238	**0.7249±0.0220**
	Sub.14	0.5245±0.0348	0.6446±0.0073	0.6265±0.0043	0.5183±0.0805	0.5309±0.0414	**0.7426±0.0039**
Sub.6	Sub.4	0.7571±0.0125	0.7739±0.0067	0.7895±0.0032	0.7055±0.0316	0.7111±0.0244	**0.8337±0.0052**
	Sub.5	0.6593±0.0417	0.6716±0.0205	0.7048±0.0226	0.6633±0.0106	0.6729±0.0231	**0.8469±0.0008**
	Sub.7	0.6963±0.0097	0.7748±0.0024	0.7112±0.0042	0.6824±0.0087	0.7726±0.0096	**0.8487±0.0108**
	Sub.12	0.3994±0.0373	0.4002±0.0267	0.4118±0.0844	0.4796±0.0271	0.4531±0.0349	**0.6254±0.0156**
	Sub.14	0.4378±0.0415	0.4625±0.0437	**0.5433±0.0525**	0.5309±0.0129	0.5171±0.0466	0.5292±0.0053
Sub.7	Sub.4	0.7043±0.0032	0.7493±0.0032	0.7545±0.0026	0.7364±0.0185	0.7317±0.0046	**0.8501±0.0049**
	Sub.5	0.8470±0.0108	0.8435±0.0018	**0.8842±0.0081**	0.7921±0.0164	0.7825±0.0114	0.8791±0.0008
	Sub.6	0.6867±0.0029	0.6880±0.0071	0.6938±0.0041	**0.7607±0.0185**	0.6891±0.0082	0.7410±0.0052
	Sub.12	0.5219±0.0125	0.5307±0.0046	0.5013±0.0070	0.4867±0.0407	0.5065±0.0241	**0.7647±0.0209**
	Sub.14	0.4979±0.0195	0.5175±0.0108	0.5382±0.0251	0.5077±0.0347	0.5139±0.0337	**0.7219±0.0140**
Sub.12	Sub.4	0.5225±0.0250	0.5806±0.0361	0.5965±0.0806	0.5663±0.0466	0.5543±0.0533	**0.7144±0.0042**
	Sub.5	0.5365±0.0341	0.5528±0.0183	0.5504±0.0034	0.5401±0.0514	0.5171±0.0305	**0.6356±0.0070**
	Sub.6	0.7116±0.0240	**0.7367±0.0079**	0.7331±0.0025	0.6783±0.0653	0.6554±0.0211	0.7123±0.0179
	Sub.7	0.6945±0.0208	0.6696±0.0167	0.6870±0.0202	0.6249±0.0249	0.6377±0.0214	**0.6995±0.0039**
	Sub.14	0.4329±0.0314	0.3729±0.0818	0.3433±0.0823	0.4154±0.0443	0.3859±0.0622	**0.4471±0.0010**
Sub.14	Sub.4	0.7375±0.0158	0.7642±0.0099	0.7537±0.0142	0.7191±0.0181	0.7097±0.0175	**0.8516±0.0027**
	Sub.5	0.6900±0.0207	0.7353±0.0424	0.7835±0.0217	0.7000±0.0081	0.7312±0.0196	**0.8185±0.0094**
	Sub.6	0.7287±0.0112	0.7279±0.0213	0.7619±0.0040	0.7223±0.0284	0.7317±0.0245	**0.7851±0.0037**
	Sub.7	0.8117±0.0294	0.7649±0.0218	0.7950±0.0152	0.7208±0.0419	0.7425±0.0288	**0.8529±0.0111**
	Sub.12	0.4240±0.0168	0.5497±0.0169	**0.5814±0.0649**	0.5388±0.0351	0.5057±0.0377	0.5067±0.0427
Average		0.6484	0.6767	0.6866	0.6429	0.6490	**0.7653**

higher than other methods. More importantly, other methods show serious negative effects in some groups due to their difficulty in solving CDB and COL, which is not present in our method.

It needs to be emphasized that the training process shows that our method has a better training curve, that is, with the increase of training epochs, the accuracy gradually increases and stabilizes around the optimal value. While the accuracy curves of other methods may fluctuate greatly, or decrease as the training process goes on, which exactly reflects their difficulty in dealing with the problem of feature confusion. Fig.4 shows an example from our experiments.

IV. CONCLUSION

In this paper, we analyzed the problem of feature confusion and summarized it as two aspects: CDB and COL. Accordingly, we designed a parallel network and proposed a novel adversarial method for the cross-subject HAR task. The experimental results on USC-HAD dataset show that our method improves the average activity recognition accuracy of the target subject by 11.69%, which is better than other general domain adaptation methods.

REFERENCES

[1] C.-T. Yen, J.-X. Liao, and Y.-K. Huang, "Human daily activity recognition performed using wearable inertial sensors combined with deep learning algorithms," *IEEE Access*, vol. 8, pp. 174105–174114, 2020.

[2] A. Barna, A. K. M. Masum, M. E. Hossain, E. H. Bahadur, and M. S. Alam, "A study on human activity recognition using gyroscope, accelerometer, temperature and humidity data," in *2019 international conference on electrical, computer and communication engineering (ecce)*, pp. 1–6, IEEE, 2019.

[3] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. De Munari, "Iot wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8553–8562, 2019.

[4] Y. Ye, Z. Huang, T. Pan, J. Li, and H. T. Shen, "Reducing bias to source samples for unsupervised domain adaptation," *Neural Networks*, 2021.

[5] R. Ding, X. Li, L. Nie, J. Li, X. Si, D. Chu, G. Liu, and D. Zhan, "Empirical study and improvement on deep transfer learning for human activity recognition," *Sensors*, vol. 19, no. 1, p. 57, 2019.

[6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[7] E. Soleimani and E. Nazerfard, "Cross-subject transfer learning in human activity recognition systems using generative adversarial networks," *Neurocomputing*, vol. 426, pp. 26–34, 2021.

[8] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International conference on machine learning*, pp. 1180–1189, PMLR, 2015.

[9] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[10] Y. Jin, X. Wang, M. Long, and J. Wang, "Minimum class confusion for versatile domain adaptation," in *European Conference on Computer Vision*, pp. 464–480, Springer, 2020.

[11] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proceedings of the 23rd ACM international conference on Multimedia*, pp. 1307–1310, 2015.

[12] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3723–3732, 2018.

[13] M. Zhang and A. A. Sawchuk, "Usc-had: a daily activity dataset for ubiquitous activity recognition using wearable sensors," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 1036–1043, 2012.

[14] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," *arXiv preprint arXiv:1705.10667*, 2017.