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

End-to-End Versatile Human Activity Recognition with Activity Image Transfer Learning

Yalan Ye[1,*], Ziqi Liu[1], Ziwei Huang[1], Tongjie Pan[1] and Zhengyi Wan[1]

***Abstract*— Transfer learning is a common solution to address cross-domain identification problems in Human Activity Recognition (HAR). Most existing approaches typically perform cross-subject transferring while ignoring transfers between different sensors or body parts, which limits the application scope of these models. Only a few approaches have been made to design a versatile HAR approach (cross-subject, cross-sensor and cross-body-part). Unfortunately, these existing approaches depend on complex handcrafted features and ignore the inequality of samples for positive transfer, which will hinder the transfer performance. In this paper, we propose a framework for versatile cross-domain activity recognition. Specifically, the proposed framework allows end-to-end implementation by exploiting adaptive features from activity image instead of extracting handcrafted features. And the framework uses a two-stage adaptation strategy consisting of pretraining stage and re-weighting stage to perform knowledge transfer. The pretraining stage ensures transferability of the source domain as well as separability of the target domain, and the re-weighting stage rebalances the contribution of the two domain samples. These two stages enhance the ability of knowledge transfer. We evaluate the performance of the proposed framework by conducting comprehensive experiments on three public HAR datasets (DSADS, OPPORTUNITY, and PAMAP2), and the experimental results demonstrate the effectiveness of our framework in versatile cross-domain HAR.**

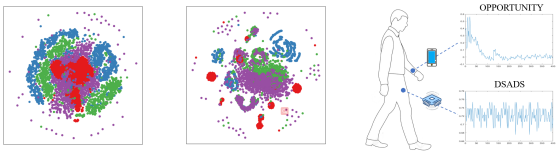
I. INTRODUCTION

In recent years, Human Activity Recognition (HAR) becomes a very attractive research area due to its potential applications such as fall detection [1] and smart home sensing [2]. Transfer Learning (TL) is wildly used in HAR to solve the cross-domain identification problems due to its capability of transferring knowledge from well-labeled source domain to unlabeled target domain. Nonetheless, existing TL approaches typically focus on cross-subject transfer while ignoring cross-sensor and cross-body-part transfers. Namely, the user needs to change the models when using another kinds of sensors or measuring on different body parts, which limits the application of the models. Thus, it is noteworthy to implement a versatile cross-domain approach in practical application scenarios.

*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, Ziqi Liu, Ziwei Huang, Tongjie Pan 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



(a) OPP (b) DSADS (c) Walking signals

Fig. 1. Comparison of signal distributions between different subjects, different sensors and different body parts. (a) and (b) show the four common activities distributions of different subjects in the OPP and DSADS datasets respectively. (c) shows two normalized walking signals on different body parts in OPP and DSADS datasets which are collected by different sensors. The units in (c) represent different types of sensors.

Implementation of such a versatile cross-domain approach is difficult since there are distinct distributions between different subjects, different sensors and different body parts. The distribution of DSADS in Fig. 1(a), which is more complicated than that of OPP in Fig. 1(b), leads to difficulty of classification. As a result, the cross-subject transfer performance on DSADS drop sharply compared with OPP. Fig. 1(c) shows that activity signals collected from different sensors and body parts follow totally different distributions, which will hinder the cross-sensor and the cross-body-part transferring performance.

Several attempts are made to solve the distribution discrepancy problem [3], [4], [5], [6], [7]. Wang et al. choose the right source domain which has the most similar properties with the target domain to alleviate the domain discrepancy [4]. Prabono et al. learn a latent representation that minimizes the discrepancy between domains by reducing statistical distance [5]. These methods extract the complex handcrafted features from original signals as the input. For example, Wang et al. obtain 27 features in both time and frequency domains, including the average value of samples, standard deviation, and so on [4]. The complexity of handcrafted features leads to the difficulty of knowledge transferring. Besides, existing methods [3] [4] [5] consider each sample has the same contribution to the positive transfer. However, some samples contain more domain-specific features that can cause negative transfer when mapped into the subspace in the same way as other samples. These two reasons result in insufficient performance of existing methods.

In this paper, we propose an end-to-end versatile cross-domain activity recognition framework referred to as Activity Image Transfer Learning (AITL). Compared to existing methods, AITL makes end-to-end implementation by extracting the adaptive features from activity image, which considers the correlation among any pair of signals instead of extracting independently from multiple time-series sensor

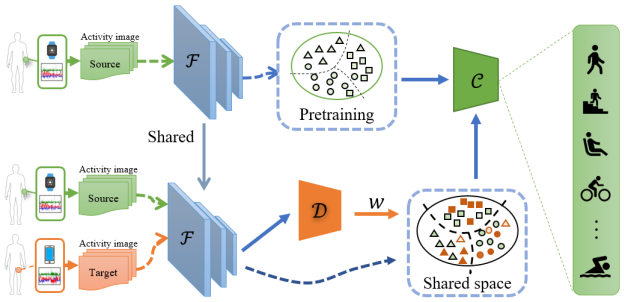


Fig. 2. The overview of the proposed AITL. Activity image are obtained from sensor signals. Firstly, AITL performs pretraining on source domain by triplet loss. Secondly, both the source and target samples are fed into domain discriminator D to be re-weighted. Finally, re-weighted samples are mapped into shared subspace and the features are classified by classifier C. The weights of samples are represented by the shade of the colors and the two generators in figure share parameters.

signals in a handcrafted way. Besides, AITL transfers knowledge by employing a two-stage adaptation strategy. Firstly, triplet loss is used in pretraining stage to ensure the transferability of source domain and separability of target domain. Secondly, both the source and target samples are re-weighted by domain discriminator to rebalance contributions of each sample. The combination of the two stages ensures that features more suitable for transfer are extracted. In this way, the transfer performance will be improved. Comprehensive experiments on three large public HAR datasets (DSADS [8], OPP [9] and PAMAP2 [10]) are conducted to evaluate the performance of AITL. The experimental results show that our method exceeds the existing state-of-the-art methods by 3% in cross-sensor recognition tasks and has a 2% improvement in cross-subject recognition tasks, and is slightly weaker than existing methods in cross-body-part recognition tasks.

The main contributions of this paper can be summarized as follows:

- We propose a versatile cross-domain HAR framework AITL, which can be used in different HAR transfer learning scenarios including cross-subject, cross-sensor and cross-body-part.
- AITL uses the activity image as the input due to the property of simple distribution. The discard of handcrafted features implements end-to-end framework and avoids the complex distribution.
- A staged adaptation strategy is designed for knowledge transfer. The transferability of source domain and separability of target domain are increased in pretraining stage. Subsequently, AITL re-weights samples and extracts common features of the re-weighted samples. The two stages ensure the transfer performance even if the distribution discrepancy is large.

II. METHOD

Fig. 2 shows the overview of AITL. The feature generator F is pretrained on source domain by triplet loss, which makes source domain features more suitable for transfer and target domain features more separable. The domain discriminator D is trained to evaluate how similar a sample is to another domain. Both source and target domain samples are weighted according to the confusion degree they cause to the discriminator. Then, the two domain samples are mapped into the shared subspace and classified by classifier C. Note that the two generators in figure share parameters.

A. Activity image for adaptive features extraction

Instead of exploiting handcrafted features, AITL extracts adaptive features from activity images. We use activity image based on 3-axis gyroscope and 3-axis linear acceleration signals. According to [11], we stack raw signals row-by-row into signal image, which allows every signal sequence has the chance to be adjacent to every other sequence. Then, 2D Discrete Fourier Transform (DFT) is applied to the stacked signal image and its magnitude is chosen as activity image.

The visualization of activity image is shown in Fig. 3. The middle of the activity image reflects the low-frequency information of the original signal, and the two sides reflect the high-frequency information. Note that different activity images differ greatly in the low-frequency information, which is the basis for the active image as input data.

(a) Walking (b) Sitting

(c) Lying (d) Standing

Fig. 3. Visualization results of activity image corresponding to four common activities of the used three datasets. It is shown that different activity images have great discrepancy in low-frequency information.

B. Pretraining for deep information exploiting

AITL fully exploits the deep information of source domain based on consideration that the source domain is well-labeled. Triplet loss is used on the source domain to make the samples within the same class more compact and increase the inter-class distance. Online generation method [12] is used to acquire all possible triplets $\mathcal{T}$. The anchor, positive and negative sample are donated as x_i^a, x_i^p and x_i^n respectively. The loss can be formulated as equation (1). After pretraining we can assume that AITL clusters each class compactly and is partially capable of transferring.

$$\mathcal{L}_{tri} = \sum_{i}^{N} (\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha). \quad (1)$$

C. Re-weighting and knowledge transferring

Existing approaches [3], [4], [5] typically consider each sample equally. However, these works ignore that not all samples are suitable for transfer. Inspired by this, we use a domain discriminator to re-weight samples via probability predictions. The sample which is more likely to confuse the discriminator have a larger weight, i.e., samples are more suitable for tranferring make greater contribution to knowledge transfer than others. We train the discriminator D by using BCE loss, where $y_i = 1$ and $y_i = 0$ indicate the ground truth of source and target samples respectively, $\hat{y}_i^d$ and $1 - \hat{y}_i^d$ are the corresponding prediction which shows how likely a sample is to belong to the source domain and

target domain respectively. By minimizing the BCE loss, D will obtain the capability to evaluate how similar a sample is to another domain. The objective is as follows:

$$\mathcal{L}_d = -y_i \log \hat{y}_i^d + (1 - y_i) \log(1 - \hat{y}_i^d). \tag{2}$$

Note that the more likely a sample is to be misjudged by the discriminator the more similar the sample is to the other domain, i.e., samples with a high probability of misclassification are suitable to exploit the shared features. As mentioned above, the misjudged probability can be measured via predictions of D. The similarity between ith source domain sample and target domain can be formulated as $1 - \hat{y}_{s_i}^d$, and $\hat{y}_{t_i}^d$ is the similarity between ith target domain sample and source domain. We calculate the weights in the exponential form of e. Considering the uncertainty of the target domain sample, We reduce and truncate target domain weights. The weights of source domain and target domain samples can be formulated as equation (3), where $w_i^s \in [1, e]$ and $w_i^t \in [0, e-1]$ represent the weights of the ith sample of source and target domains respectively.

$$\begin{cases} w_i^s = e^{1-\hat{y}_{s_i}^d} \\ w_i^t = max\left\{-1 + e^{\hat{y}_{t_i}^d},\ 0\right\}. \end{cases} \tag{3}$$

After re-weighting samples, we conduct supervised learning on source domain by minimizing the weighted cross entropy loss which empowers feature generator F to exploit the common features of the two domains. Besides, we consider increase classification confidence by minimizing the weighted entropy loss on target samples. The classification prediction of source sample x_i^s is denoted as $\hat{y}_i^s = C(F(x_i^s))$ and the classification probability of target sample is formulated as $\hat{y}_i^t = C(F(x_i^t))$. The weighted loss is denoted as follows:

$$\mathcal{L}_w = -[\frac{1}{|D_s|} \sum_{x_s \in D_s} (w_i^s y_i^s \log \hat{y}_i^s) + \frac{1}{|D_t|} \sum_{x_t \in D_t} (w_i^t \hat{y}_i^t \log \hat{y}_i^t)]. \tag{4}$$

While the value of weighted entropy tends to zero with minimization, which shows the necessity of entropy loss for the minimization of the target risk. However, entropy minimization is necessary but not sufficient. AITL employs the entropy of $P^N = (p^1, p^2, \ldots, p^N)$ as regularization, where p^n and the entropy of P^N is formulated as equation (5) and (6) respectively.

$$p^n = \frac{1}{|D_t|} \sum_{x_t \in D_t} P(y_{x_t} = n | x_t), \tag{5}$$

$$\mathcal{L}_e = -\sum_{n=1}^{N} (p_i^n \log p_i^n). \tag{6}$$

The complete training objective can be formulated by the combination of $\mathcal{L}_w$ and $\mathcal{L}_e$, where $\lambda \geq 0$ is weighing factor. Optimized by equation (7), AITL can perform knowledge transfer between different subjects, sensors and body parts.

$$\mathcal{L} = \mathcal{L}_w + \lambda \mathcal{L}_e. \tag{7}$$

III. EXPERIMENTS

We evaluate the adaptation performance of AITL on three large public datasets DSADS , OPP and PAMAP2. The brief introduction of these three datasets are given in Table I. Each dataset is collected by different types of sensors, and the data is acquired on several different body parts of subjects. Note that the data of each body part has the same proportion in the dataset.

TABLE I
INFORMATION OF THE THREE HAR DATASETS.

Dataset	Subject	Activity	Sensor	Body parts
DSADS	8	19	Xsens MTx units	Right Arm(RA), Left Arm(LA), Right Leg(RL), Left Leg(LL), Tarso(T)
OPP	4	4	MARG sensors	Right Upper Arm(RUA), Right Lower Arm(RLA), Left Upper Arm(LLA), Left Lower Arm(LLA), Back (B)
PAMAP2	9	18	IMU	Hand(H), Chest(C), Ankle (A)

A. Experimental Setup

In order to evaluate the performance of the proposed AITL, we conduct comprehensive experiments on the following three HAR scenarios: (I) cross-subject transfer within the same dataset and similar body part. (II) cross-body-part transfer within the same subject and different body part. (III) cross-sensor transfer between different subjects and similar body part collected by different types of sensors. The notation $A \rightarrow B$ is used to denote knowledge transferring from labeled domain A to unlabeled domain B. For scenarios (I) and (II), we use all the classes of each dataset; for scenario (III), we extract 4 common classes for each dataset (i.e., walking, sitting, lying, and standing).

B. Evaluation of AITL

The classification accuracy of the target domain are listed in TABLE II. We compare AITL with other existing approaches. The results for PCA [13], KPCA [13], TCA [14], GFK [15] and TKL [16] are derived from [4]. It is shown that AITL outperforms the existing state-of-the-art methods by 3% in cross-sensor tasks and 2% in cross-subject tasks, and is slightly weaker than existing methods in cross-body-part tasks. Weakness of AITL in the cross-body-part task may be caused by the large differences in data from dissimilar body parts, as AITL focuses on further improving the quality of common features with similar data.

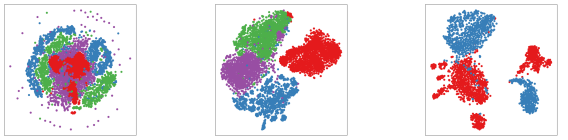
We further evaluate our framework by visualizing the adaptation results of the most difficult scenario (III) tasks since the tasks of (III) transfer between totally dissimilar body parts. Fig. 4(a) and Fig. 4(d) show the handcrafted features of OPP and DSADS datasets follow complex distributions. In contrast, the corresponding distributions of activity image shown in Fig. 4(b) and Fig. 4(e) are simple, and the features of activity image are compactly clustered, which demonstrate that activity image is more suitable as input than handcrafted features.

TABLE II

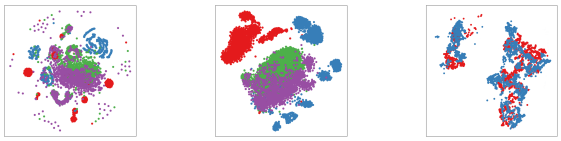
COMPARISON OF THE CLASSIFICATION ACCURACY

Scenario	Dataset	Task	PCA	KPCA	TCA	GFK	TKL	STL	TNNAR	Asura	**Ours**
I	DSADS	Sub(RA)->Sub(LA)	59.91	62.17	66.15	71.07	54.10	71.04	75.89	73.50	**80.38**
		Sub(RL)->Sub(LL)	69.46	70.92	75.06	79.71	61.63	81.60	86.76	70.89	**89.15**
	OPP	Sub(RUA)->Sub(LUA)	76.12	65.64	76.88	74.62	66.81	83.96	87.43	**92.67**	89.56
		Sub(RLA)->Sub(LLA)	62.17	66.48	60.60	74.62	66.82	83.93	86.29	89.86	**92.21**
II	DSADS	RA->T	38.89	30.20	39.41	44.19	32.72	45.61	50.22	58.33	**58.74**
	PAMAP2	H->C	34.97	24.44	34.86	36.24	35.67	43.47	46.32	45.93	**49.95**
	OPP	RLA->T	59.10	46.99	55.43	48.49	47.66	56.88	59.58	**79.95**	62.45
		RUA->T	67.95	54.52	67.50	66.14	60.49	75.15	75.75	**89.07**	84.56
III	PAMAP2->OPP	IMU(C)->MARG(B)	32.80	43.78	39.02	27.64	35.64	40.10	45.62	50.04	**53.72**
	DSADS->PAMAP2	XMT(T)->IMU(C)	23.19	17.95	23.66	19.39	21.65	37.83	39.21	36.24	**40.23**
	OPP->DSADS	MARG(B)->XMT(T)	44.30	49.35	46.91	48.07	52.79	55.45	57.97	60.31	**64.65**
Average			51.71	48.40	53.23	53.69	48.73	61.37	64.64	67.90	**69.60**

[1] (I) cross-subject and cross-body-part (similar) transfer. (II) cross-body-part (dissimilar) transfer. (III) cross-sensor and cross-body-part (similar) transfer.
[2] Xsens MTx units (XMT), MARG sensors (MARG)



(a) Original OPP (b) Activity images of OPP dataset (c) Source only



(d) Original DSADS (e) Activity images of DSADS dataset (f) Ours

Fig. 4. Effectiveness demonstration of AITL. (a) and (d) show the handcrafted feature distributions of the four activities common to OPP and DSADS datasets. (b) and (e) show the corresponding activity image distributions. (c) and (f) show the results of source only model and our staged adaptation respectively, where blue points indicate the source domain and red points indicate the target domain.

The domain adaptation results are visualized in Fig. 4(c) and Fig. 4(f), which show the performance of source only model and AITL. The OPP and DSADS datasets are used as source and target domain respectively. For the source only model, we use the same network architecture as used in AITL. Fig. 4(c) shows that distributions of source domain and target domain mismatch totally. The distributions of the source domain and the target domain are aligned in Fig. 4(f), which demonstrates the effectiveness of AITL.

IV. CONCLUSIONS

In this paper, we propose an end-to-end versatile cross-domain HAR framework AITL, which is capable of cross-subject, cross-sensor and cross-body-part knowledge transfers. The proposed framework extracts adaptive features from activity image due to simple distribution, thus avoiding the extraction of handcrafted features and implementing end-to-end HAR framework. The combination of pretraining stage and re-weighting stage enables AITL to transfer activity knowledge even if the distribution discrepancy is large. The experimental results demonstrate the effectiveness of AITL in versatile cross-domain HAR.

REFERENCES

[1] S. A. W. Talha, A. Fleury, and S. Lecoeuche, "Hierarchical classification scheme for real-time recognition of physical activities and postural transitions using smartphone inertial sensors," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1243–1246, IEEE, 2019.

[2] L. Jing and Z. Cheng, "Recognition of daily routines and accidental event with multipoint wearable inertial sensing for seniors home care," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2324–2389, IEEE, 2017.

[3] J. Wang, Y. Chen, L. Hu, X. Peng, and S. Y. Philip, "Stratified transfer learning for cross-domain activity recognition," in *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–10, IEEE, 2018.

[4] J. Wang, V. W. Zheng, Y. Chen, and M. Huang, "Deep transfer learning for cross-domain activity recognition," in *proceedings of the 3rd International Conference on Crowd Science and Engineering*, pp. 1–8, 2018.

[5] A. G. Prabono, B. N. Yahya, and S.-L. Lee, "Atypical sample regularizer autoencoder for cross-domain human activity recognition," *Information Systems Frontiers*, vol. 23, no. 1, pp. 71–80, 2021.

[6] Y. Ye, Y. He, T. Pan, J. Li, and H. T. Shen, "Alleviating domain shift via discriminative learning for generalized zero-shot learning," *IEEE Transactions on Multimedia*, 2021.

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

[8] B. Barshan and M. C. Yüksek, "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units," *The Computer Journal*, vol. 57, no. 11, pp. 1649–1667, 2014.

[9] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. d. R. Millán, and D. Roggen, "The opportunity challenge: A benchmark database for on-body sensor-based activity recognition," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 2033–2042, 2013.

[10] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *2012 16th International Symposium on Wearable Computers*, pp. 108–109, IEEE, 2012.

[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] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

[13] I. K. Fodor, "A survey of dimension reduction techniques," tech. rep., Citeseer, 2002.

[14] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.

[15] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *2012 IEEE conference on computer vision and pattern recognition*, pp. 2066–2073, IEEE, 2012.

[16] M. Long, J. Wang, J. Sun, and S. Y. Philip, "Domain invariant transfer kernel learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1519–1532, 2014.