

## Semisupervised Human Activity Recognition With Radar Micro-Doppler Signatures

Li, Xinyu; He, Yuan; Fioranelli, Francesco; Jing, Xiaojun

**DOI**

[10.1109/TGRS.2021.3090106](https://doi.org/10.1109/TGRS.2021.3090106)

**Publication date**

2022

**Document Version**

Accepted author manuscript

**Published in**

IEEE Transactions on Geoscience and Remote Sensing

**Citation (APA)**

Li, X., He, Y., Fioranelli, F., & Jing, X. (2022). Semisupervised Human Activity Recognition With Radar Micro-Doppler Signatures. *IEEE Transactions on Geoscience and Remote Sensing*, 60.  
<https://doi.org/10.1109/TGRS.2021.3090106>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# Semisupervised Human Activity Recognition With Radar Micro-Doppler Signatures

Xinyu Li<sup>✉</sup>, *Student Member, IEEE*, Yuan He<sup>✉</sup>, *Member, IEEE*,  
 Francesco Fioranelli<sup>✉</sup>, *Senior Member, IEEE*, and Xiaojun Jing<sup>✉</sup>, *Member, IEEE*

**Abstract**—Human activity recognition (HAR) plays a vital role in many applications, such as surveillance, in-home monitoring, and health care. Portable radar sensor has been increasingly used in HAR systems in combination with deep learning (DL). However, it is both difficult and time-consuming to obtain a large-scale radar dataset with reliable labels. Insufficient labeled data often limit the generalization of DL models. As a result, the performance of DL models will drop when being applied to a new scenario. In this sense, only labeling a small portion of data in the large-scale radar dataset is more feasible. In this article, we propose a semisupervised transfer learning (TL) algorithm, “joint domain and semantic transfer learning (JDS-TL),” for radar-based HAR, which is composed of two modules: unsupervised domain adaptation (DA) and supervised semantic transfer. By employing a sparsely labeled dataset to train the HAR model, the proposed method alleviates the need of labeling a significantly large number of radar signals. We adopt a public radar micro-Doppler spectrogram dataset including six human activities to evaluate JDS-TL. Experiments show that the proposed JDS-TL is able to recognize the six activities with an average accuracy of 87.6% when there are only 10% instances labeled in the training dataset. Ablation analysis also demonstrates the efficiency of the DA and the semantic transfer modules.

**Index Terms**—Human activity recognition (HAR), radar micro-Doppler (MD) effect, semisupervised learning, transfer learning (TL).

## I. INTRODUCTION

WITH aging population worldwide, eldercare and health-care by monitoring inhabitants and their daily activities are more and more necessary [1]–[4]. Human activity recognition (HAR), which can help smart systems take further actions quickly to improve the well-being of people in terms of comfort and health, has been a vital underpinning for numerous applications, such as surveillance, assisted living and smart homes [5]–[8].

Manuscript received December 30, 2020; revised May 4, 2021; accepted June 15, 2021. This work was supported in part by the National Nature Science Foundation of China under Grant 61901049 and in part by the Beijing University of Posts and Telecommunications (BUPT) Excellent Ph.D. Students Foundation under Grant CX2020208. (Corresponding author: Yuan He.)

Xinyu Li, Yuan He, and Xiaojun Jing are with the Key Laboratory of Trustworthy Distributed Computing and Service (BUPT), Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: lixinyu@bupt.edu.cn; yuanhe@bupt.edu.cn; jxiaojun@bupt.edu.cn).

Francesco Fioranelli is with the Department of Microelectronics, Delft University of Technology, 2628 Delft, The Netherlands (e-mail: f.fioranelli@tudelft.nl).

Digital Object Identifier 10.1109/TGRS.2021.3090106

Transfer learning (TL) is often adopted for classification when there is little or no labeled data in the training dataset. By introducing a preexisting dataset that has different but related data distribution from the target training dataset, TL can address the problem that the target dataset cannot provide sufficient label information. In this case, prior knowledge in the source dataset is extracted and transferred to classify data in the target dataset. Currently, depending on whether the target datasets are labeling or not, two kinds of TL methods, i.e., the supervised and unsupervised TL methods, are proposed for HAR. The supervised TL [9]–[12] utilizes labeled radar data in the target dataset to transfer the source prior knowledge. Such an approach can achieve good performance when sufficient labeled data are available for each class. On the contrary, the unsupervised TL [13], [14], based on domain adaptation (DA) with unlabeled target data, is employed to learn domain-invariant feature representation. However, due to the lack of label information, the performance of the unsupervised methods is generally not as good as the supervised ones.

On the other hand, compared with labeled radar datasets, using sparsely labeled datasets, where only a small number of instances are labeled, is generally more applicable for practical applications. It is because labeling a vast amount of instances, especially radar signals, is time-consuming and expensive due to the much human effort required. It is also difficult to get and reliably label a lot of radar data. Furthermore, in sparsely labeled datasets, it is not required that the number of labeled instances in each class be roughly the same.

In this article, we propose a semisupervised TL algorithm, “joint domain and semantic transfer learning (JDS-TL),” to train a HAR model using a sparsely labeled radar dataset. The proposed JDS-TL is composed of two modules: unsupervised adversarial domain adaptation (ADA) and supervised semantic transfer. Specifically, we utilize the unsupervised ADA [15] to transfer the source knowledge from a preexisting labeled dataset (*source domain*) to the sparsely labeled dataset (*target domain*). ADA can mitigate the detrimental effects of the domain shift and the dataset bias by mapping the source and target data into a common feature space.

Furthermore, a supervised semantic transfer method is proposed. Knowledge distillation via soft labels [16] is adopted in the semantic transfer to learn interclass and intraclass information from the source domain. Since there are sufficient interclass correlations in the soft labels, more useful

information can be learned under the supervision of soft labels.

The contributions of this article are summarized as follows.

- 1) We propose a semisupervised TL method, *JDS-TL*, for radar-based HAR. The method can recognize human activities under realistic situations where the only available dataset is sparsely labeled.
- 2) There are two modules in *JDS-TL*, i.e., the unsupervised ADA module and the supervised semantic transfer module. We propose the multilayer ADA to learn domain-invariant feature representation in an unsupervised manner and develop the semantic transfer via knowledge distillation to transfer the interclass and intraclass semantic information.
- 3) To evaluate the performance of *JDS-TL* algorithm, we perform experiments on a public radar-based HAR dataset [17] including six daily human activities. Experimental results show that *JDS-TL* achieves a recognition accuracy of 87.6% for HAR, even when there are merely 10% spectrograms labeled in the target dataset. Ablation analysis also demonstrates the efficiency of *JDS-TL* for training a HAR model with the sparsely labeled dataset.

The remainder of this article is organized as follows. Section II reviews the recent research on HAR, unsupervised domain adaptation (UDA), and supervised semantic transfer. In Section III, the proposed *JDS-TL* algorithm is presented in detail. The datasets used to evaluate *JDS-TL* are described, and experiment implementation is given in Section IV. Section V presents the experimental results and some discussions. Finally, a conclusion is drawn in Section VI.

## II. RELATED WORKS

### A. Human Activity Recognition

Gurbuz and Amin [18] introduced deep learning (DL)-based data-driven approach for motion classification in indoor monitoring areas. Le Kernec *et al.* [19] proposed radar signal processing approaches for assisted living through three typical applications, i.e., human daily activity recognition, respiratory disorders, and sleep stages classification. Singh *et al.* [20] investigated the performance of several sensors, such as accelerometer, pressure sensor, radar, camera, and infusion, for fall detection in daily life. Chaccour *et al.* [21] summarized the existing fall-related systems and divided them into three categories, i.e., the wearable method, the non-wearable method, and the fusion method, according to the sensor deployment. However, almost all these methods, either supervised or unsupervised, do not make full use of sparsely labeled radar datasets, which are realistically available. In this article, we propose the semisupervised TL method especially for sparsely labeled datasets.

### B. Unsupervised Deep DA

It is common that there is a distribution shift between the data for training a DL model and the data for testing the model. Such distribution shift, typically referred to as the domain shift or the dataset bias [15], can degrade the performance of a

trained classifier at the test stage. In radar data for HAR, due to the environmental factors (e.g., different rooms, furniture) and human individuals' differences, the distribution shift often occurs, degrading the performance of a trained classifier when being applied to a new radar dataset.

UDA is an unsupervised TL method that mitigates domain shifts. Since the unlabeled data are much easier to acquire on radar, UDA is more suitable for radar-based HAR with unlabeled data. Several UDA methods have been proposed for radar-based HAR. By utilizing the motion capture database as the source dataset for knowledge transferring, Lang *et al.* [14] proposed a UDA method to learn the domain-invariant features to classify the unlabeled measured radar data. Du *et al.* [13] utilized an unsupervised adversarial domain adaption method to reduce the domain discrepancy between the simulated radar spectrogram dataset and the measured spectrogram dataset. Chen *et al.* [22] proposed two adaptation networks that utilized DA to eliminate the impact of aspect angle on HAR with micro-Doppler (MD) spectrograms.

In this article, to take advantage of the unlabeled data in the sparsely labeled dataset, we adopt UDA as a component of *JDS-TL* and propose a multilayer discriminator model. The proposed discriminator is different from the general domain discriminator that distinguishes the fake vs real images with the output from a single layer of the generator. At each layer of the proposed discriminator, information is accumulated from both the output by a previous layer of the discriminator and from the output by a specific layer of the generator. The proposed discriminator allows deeper alignment of feature representations and thus can achieve stable adversarial learning of UDA.

### C. Knowledge Distillation via Soft Labels

Hinton *et al.* [16] introduced knowledge distillation by transferring the rich knowledge of a task-related well-trained model to other deep models. In deep discriminative networks, there is more information in logits than that in hard labels [23]. For example, for a classification task, the intraclass correlation cannot be indicated in hard labels, which are often represented as one-hot codes. However, there is sufficient similarity information between classes in the logits. Hence, soft label [24] is proposed as a knowledge distillation approach by processing the logits and retaining more interclass information. Note that though more interclass similarity information is retained, the soft label is still class-discriminative.

Romero *et al.* [25] utilized soft output labels and intermediate representations learned by the teacher network as hints to supervise the training of a student network. Yim *et al.* [26] proposed that the feature flow between layers can be distilled and transferred to another deep network. Yang *et al.* [27] presented an extra loss term to train the student network, which is able to make the teacher provide a milder supervision signal and transfer more interclass similarity information to the student network.

Knowledge distillation is often utilized to train a smaller model to meet limited resource environment via a teacher-student structure. In this article, we integrate

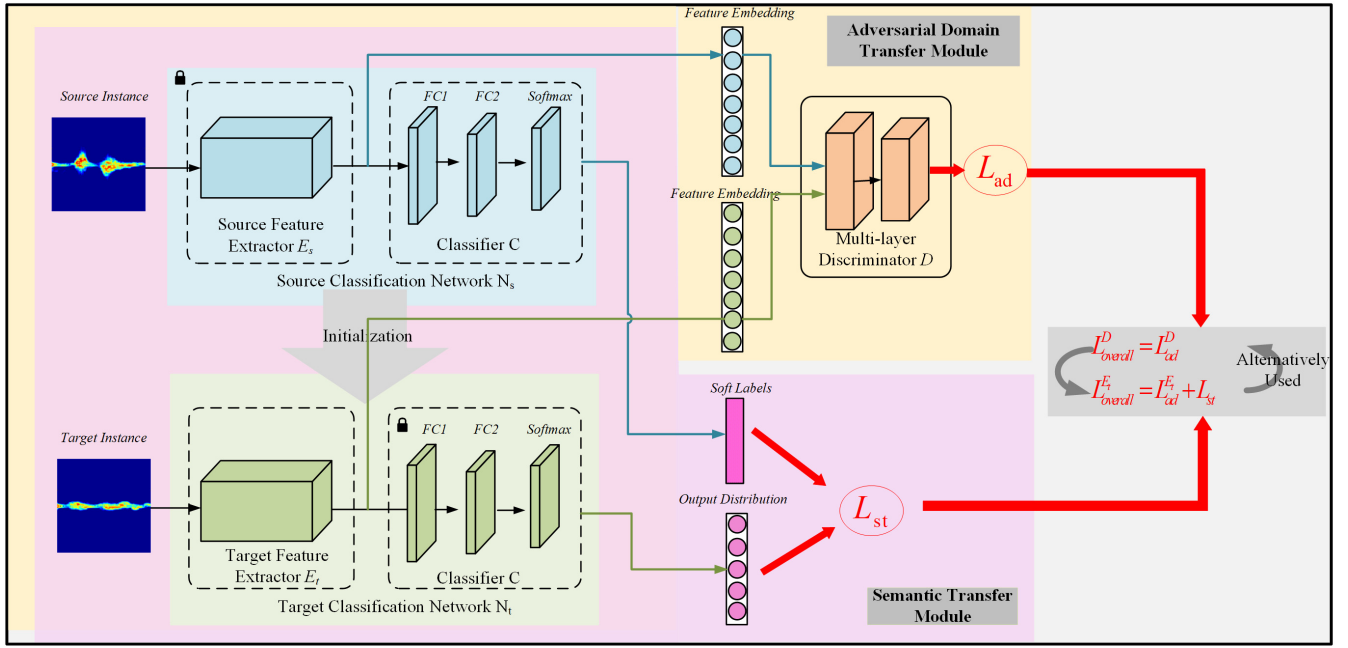


Fig. 1. Pipeline of the proposed semisupervised *JDS-TL* method. The proposed *JDS-TL* method is composed of two modules: the adversarial domain transfer module and the semantic transfer module. The adversarial domain module is able to extract domain-invariant feature representation in an unsupervised way, while the semantic transfer module transfers both the intraclass information and the interclass information from the source domain. The overall training loss  $L_{\text{overall}}$  is the sum of the adversarial training loss and the semantic transfer loss. By alternatively minimizing the losses  $L_{\text{overall}}^E$  and  $L_{\text{overall}}^D$ , the target feature extractor  $E_t$  and the discriminator  $D$  are optimized. Note that the parameters of the parts that are marked with a lock are not updated during optimization.

knowledge distillation to the semisupervised *JDS-TL* and transfer both class-discriminative information and the inter-class similarity from the source data for the target HAR task. Furthermore, the distilled interclass similarity knowledge can potentially lower the risk of overfitting.

### III. JOINT DOMAIN AND SEMANTIC TRANSFER

In this section, we propose the *JDS-TL* method and describe its components in detail.

#### A. Problem Definition

Considering a sparsely labeled radar spectrogram dataset  $X_t$  for the HAR task. The dataset consists of two parts: a small labeled subset  $X_{lt} = \{(\mathbf{x}_{lt}^j, \mathbf{y}_{lt}^j), j = 1, \dots, k\}$  (where  $\mathbf{y}_{lt}^j$  denotes the  $j$ th label corresponding to the  $j$ th instance  $\mathbf{x}_{lt}^j$ ) and an unlabeled subset  $X_{ut} = \{\mathbf{x}_{ut}^p, p = 1, \dots, m\}$ . Our aim is to train a DL model to classify the spectrograms and recognize the corresponding activities in  $X_t$ . Since only a small portion of data is labeled, it generally cannot achieve satisfactory performance if the DL model is trained with the limited labeled data from scratch. In this circumstance, we employ TL to extract useful features from the sparsely labeled dataset. Specifically, we introduce the source domain  $D_s$ , which has a set of labeled source instances  $X_s = \{(\mathbf{x}_s^i, \mathbf{y}_s^i), i = 1, 2, \dots, n\}$ , as an auxiliary, where  $\mathbf{y}_s^i$  is the label of the  $i$ th instance  $\mathbf{x}_s^i$ . The sparsely labeled dataset is then used as the target domain  $D_t$ . Although the source dataset  $X_s$  and the target dataset  $X_t$  have the same set of activities to classify, their distributions, i.e.,  $P_s(X_s) P_t(X_t)$ , are different.

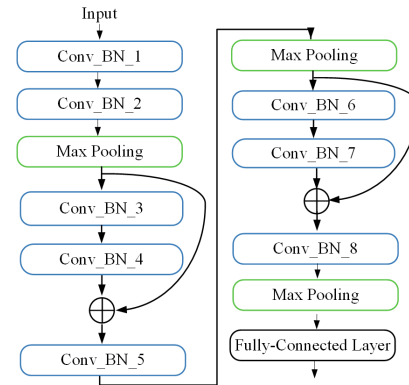


Fig. 2. Structure of the feature extractor  $E$ . The layer “Conv\_BN\_ $n$ ” represents the  $n$ th layer where there is a convolutional operation with a  $3 \times 3$  kernel and a batch normalization operation.

In this article, a target classification network  $N_t$ , including a feature extractor  $E_t$  and a classifier  $C_t$ , is trained to classify radar spectrograms in the target domain. Fig. 1 shows the flowchart of the proposed *JDS-TL* method. The two modules in *JDS-TL*, i.e., multilayer adversarial domain transfer and semantic transfer, are discussed in the following.

#### B. Multilayer Adversarial Domain Transfer

1) *Source/Target Feature Extractor*: The source feature extractor  $E_s$  and the target feature extractor  $E_t$  share the same architecture  $E$ , as shown in Fig. 2. A residual mechanism is adopted in this architecture due to its efficiency and competitiveness. With the proposed architecture, the feature



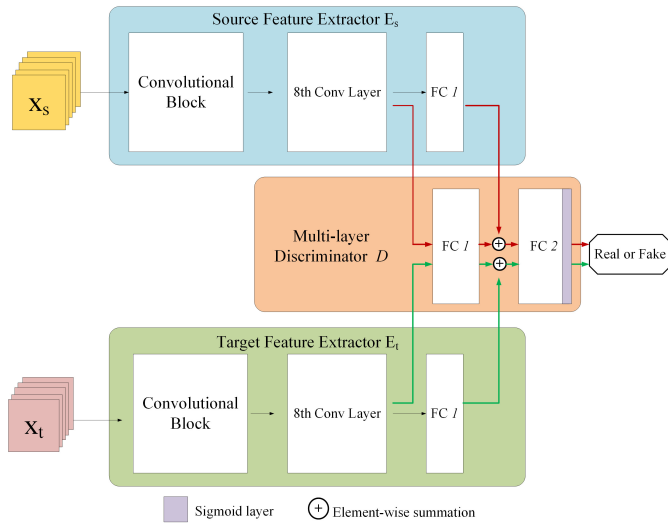


Fig. 3. Illustration of the proposed multilayer adversarial domain transfer method. To show the proposed domain transfer mechanism clearly and omit irrelevant parts, the convolution structure in the feature extractors is not presented in detail but represented by a block. The source feature extractor  $E_s$  is pretrained with the source dataset, and its parameters are frozen during the adversarial TL. The target feature extractor  $E_t$  shares the same structure as  $E_s$ , and is initialized by  $E_s$ . It is noted that the structures of the first two layers of the discriminator  $D$  are the same as those of the last two layers of  $E_t$ , so that the element-wise summation can be performed.

extractor transforms the original radar spectrogram into a feature embedding vector  $\mathbf{v} \in \mathbb{R}^{1 \times N}$ . The source feature extractor  $E_s$  is pretrained with the source dataset, and its parameters are frozen during the adversarial TL.

2) *Multilayer Discriminator*: In adversarial domain transfer, a multilayer discriminator  $D$  is proposed, whose structure is shown in Fig. 3. The multilayer discriminator  $D$  is composed of two dense layers followed by a sigmoid layer. At each layer of the proposed  $D$ , information is accumulated from both the output by the previous layer of  $D$ , and from the output by a specific layer of the feature extractor. Thus, the output of each discriminator layer can be written as

$$\mathbf{d}_m = D_m \left( \sigma \left( \gamma \mathbf{d}_{m-1} \oplus \mathbf{v}_m(x) \right) \right) \quad (1)$$

where  $D_m$  is the structure of layer  $m$  of the discriminator  $D$ ;  $\sigma(\cdot)$  represents the activation function;  $\gamma \leq 1$  is the decay factor;  $\oplus$  is element-wise summation of the feature maps;  $\mathbf{v}_m$  is the output feature representation of layer  $m$  of  $E$ ; and  $x$  is the input instance from either source domain or target domain. The multilayer discriminator allows for deeper alignment of the source and target representations. This alignment is able to improve target classification performance as well as achieve more stable adversarial learning, which will be further demonstrated in the experiment.

Furthermore, since the general adversarial loss has issues of instability [28], we adopt the Wasserstein adversarial training loss [29], which utilizes Wasserstein distance instead of the Jensen–Shannon divergence to measure the similarity between the source representation and the target representation. In particular, there are two changes compared to the conventional adversarial losses.

We first change the general adversarial losses by dropping the log function. The general adversarial loss function  $L_{ad}$  is

$$\begin{aligned} \min_{\theta_D} L_{ad}^D(\mathbf{X}_s, \mathbf{X}_{ut}, \theta_{E_s}, \theta_{E_t}) \\ = -\mathbb{E}_{\mathbf{x}_s \sim \mathbf{X}_s} [\log(D(E_s(\mathbf{x}_s)))] \\ - \mathbb{E}_{\mathbf{x}_{ut} \sim \mathbf{X}_{ut}} [\log(1 - D(E_t(\mathbf{x}_{ut})))] \end{aligned} \quad (2)$$

$$\begin{aligned} \min_{\theta_{E_t}} L_{ad}^{E_t}(\mathbf{X}_s, \mathbf{X}_{ut}, \theta_D) \\ = -\mathbb{E}_{\mathbf{x}_{ut} \sim \mathbf{X}_{ut}} [\log(D(E_t(\mathbf{x}_{ut})))] \end{aligned} \quad (3)$$

With the first change, the improved loss function  $L_{ad}$  is expressed as

$$\begin{aligned} \min_{\theta_D} L_{ad}^D(\mathbf{X}_s, \mathbf{X}_{ut}, \theta_{E_s}, \theta_{E_t}) &= -\mathbb{E}_{\mathbf{x}_s \sim \mathbf{X}_s} [D(E_s(\mathbf{x}_s))] \\ &\quad - \mathbb{E}_{\mathbf{x}_{ut} \sim \mathbf{X}_{ut}} [1 - D(E_t(\mathbf{x}_{ut}))]; \end{aligned} \quad (4)$$

$$\min_{\theta_{E_t}} L_{ad}^{E_t}(\mathbf{X}_s, \mathbf{X}_{ut}, \theta_D) = -\mathbb{E}_{\mathbf{x}_{ut} \sim \mathbf{X}_{ut}} [D(E_t(\mathbf{x}_{ut}))]. \quad (5)$$

Another change is that the weights of  $D$  are clipped and constrained within a bounded range to make the training process converge faster.

During training, the target feature extractor (generator)  $E_t$  and the discriminator  $D$  are optimized alternatively by minimizing  $L_{ad}^{E_t}$  and  $L_{ad}^D$ , respectively.  $E_t$  is used to output target representations that confuse  $D$ , while  $D$  is employed to discriminate the output target representations from the source representations.

### C. Semantic Transfer via Soft Labels

The second module of *JDS-TL* is the supervised semantic transfer, which can learn effective feature representation with labeled target data.

For a classification model, the output of *Softmax* is a  $K$ -dimensional vector, where  $K$  is the number of classes. The value of the  $k$ th dimension indicates the probability of the input instance belonging to the  $k$ th class. Furthermore, for a trained model, the intraclass similarity can also be learned from the output probability [24]. If there are two classes similar to each other, there is often similar semantic information between them. However, according to [24], when a classification model is trained with the supervision of hard labels, the *Softmax* layer will output a very “peaked” distribution, which hides semantic information about interclass similarity. To address this issue, soft labels are adopted to obtain more interclass semantic information.

The proposed semantic transfer algorithm is illustrated in Fig. 4. Specifically, we first define the soft label  $l^k$  as the average output of *Softmax* activations when the source instances that belong to class  $k$  are input. As shown in Fig. 5, the soft label  $l^k$  can be expressed as

$$l^k = \frac{1}{n} \sum_{i=1}^n l_i^k \quad (6)$$

where  $l_i^k$  denotes the *Softmax* output of the trained  $N_s$  corresponding to the  $i$ th source instance  $x_k^i$  in Class  $k$ .

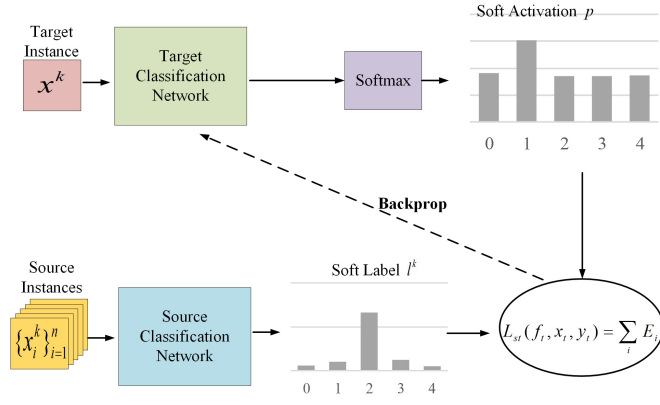


Fig. 4. Illustration of the proposed semantic transfer scheme. When a target labeled instance  $x^k$  belonging to Class  $k$  is input to the target classification network  $N_t$ , the semantic transfer loss  $L_{st}$  is obtained by calculating the information entropy  $E$  between the target soft activation  $p$  and the soft label  $l^k$ .  $l^k$  is calculated with the output of the trained source network  $N_s$ . Then, the target classification network is optimized via gradient back propagation.

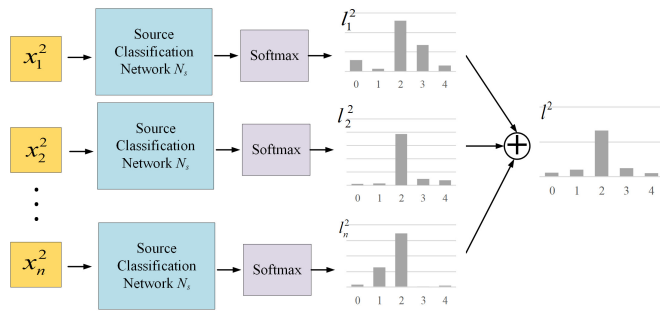


Fig. 5. Algorithm to obtain the soft label  $l^k$  for Class  $k$ . Specifically, taking the prototypical label  $l^2$  for the second class for example, it is calculated by averaging the Softmax outputs of all the source instances  $\{x_1^2, x_2^2, x_3^2, \dots, x_n^2\}$  that belong to Class 2. The source classification network  $N_s$ , which is composed of the source feature extractor  $E_s$  and the source classifier  $C_s$ , is pretrained with the labeled source dataset.

It is noted that before the calculation, the source classification network  $N_s$ , as shown in Fig. 1, is pretrained with the labeled source dataset in a supervised manner. Then, the quantities  $l^k$  are generated by the trained  $N_s$ . In this way, the soft labels of the labeled target instances [24], [25] can be obtained with the source dataset. Then, the soft labels are used to learn class-discriminative semantic information for each class in the target domain.

When a labeled target instance  $x^k$  belonging to Class  $k$  is input to  $N_t$ , the target soft activation of  $x^k$ , which is output by  $N_t$ , is given by

$$p = \text{Softmax}(\theta_t^C E_t(x^k)/\tau) \quad (7)$$

where  $\theta_t^C$  denotes the parameters of  $C_t$  and  $\tau$  is the temperature parameter.

As illustrated in Fig. 6, with diverse values of  $\tau$ , the amount of the retained intraclass semantic information varies. When the same probability distribution is output, most intraclass semantic information is retained in the input distribution when  $\tau$  equals 0.5. However, the input distribution in (a) is less discriminative than the distributions in (b) and (c), which

would increase the difficulty of classifying the corresponding instance. Therefore, setting an appropriate value of  $\tau$  can make the classification task easy while retaining sufficient intraclass semantic information. In this article,  $\tau$  is set to 0.8.

With the target soft activation  $p$  and the soft label  $l^k$ , the semantic transfer loss  $L_{st}$  can be obtained by calculating the information entropy  $E$  between  $l^k$  and  $p$ , as given by

$$\min_{\theta_{E_t}} L_{st}(X_{lt}, Y_{lt}) = - \sum_{i=1}^l \mathbb{E}_i = - \sum_{i=1}^l l_i^{y_i} \log(p^i) \quad (8)$$

where  $\mathbb{E}_i$  denotes the information entropy between the target soft activation  $p^i$  and the corresponding soft label  $l_i^{y_i}$ .

By minimizing  $L_{st}$ ,  $p^i$  will approach  $l_i^{y_i}$ , and the target classification network can be optimized via gradient backpropagation. It is noted that  $\tau$  is set to 0.8 only during the training process. When the trained model  $E_t$  is applied for testing, the value of  $\tau$  in  $N_t$  is 1.0.

In the proposed semantic transfer process, the source knowledge is transferred to the target domain by using soft labels obtained from  $N_s$ . In this way, discriminative information and intraclass correlations are extracted.

#### Algorithm 1 Training Process of JDS-TL

**Input:** A sparsely labeled dataset  $X_t$ , a labeled dataset  $X_s$

**Output:**  $\theta_{E_s} = [w_{s1}, w_{s2}, \dots, w_{sn}, b_{s1}, b_{s2}, \dots, b_{sn}]$ ,  
 $\theta_{E_t} = [w_{t1}, w_{t2}, \dots, w_{tn}, b_{t1}, b_{t2}, \dots, b_{tn}]$ ,  
 $\theta_D = [w_{d1}, w_{d2}, \dots, w_{dn}, b_{d1}, b_{d2}, \dots, b_{dn}]$

- 1 Train the source feature extractor  $E_s$  and the source classifier  $C_s$  by minimizing  $L_{sc}$  in 9;
- 2 Initialize the target feature extractor  $E_t$  with the optimized parameters of  $E_s$ ;
- 3 **while not converge do**
- 4     Update the parameters  $\theta_{E_t} = [w_{t1}, w_{t2}, \dots, w_{tn}, b_{t1}, b_{t2}, \dots, b_{tn}]$  of  $E_s$  by minimizing the loss function  $L_{overall}^{E_t} = L_{ad}^{E_t} + \alpha L_{st}$  in 13, while  $\theta_{E_s}$  and  $\theta_D$  are fixed;
- 5     Update the parameters  $\theta_D = [w_{d1}, w_{d2}, \dots, w_{dn}, b_{d1}, b_{d2}, \dots, b_{dn}]$  of  $D$  by minimizing the loss function  $L_{overall}^D = L_{ad}^D + \alpha L_{st}$  in 12, while  $\theta_{E_s}$  and  $\theta_{E_t}$  are fixed;
- 6 **end**

#### D. Training Process of JDS-TL

1) *Training Source Classification Network  $N_s$ :* The source classification network  $N_s$  (plotted as blue blocks in Fig. 1), which consists of a source feature extractor  $E_s$  and a source classifier  $C_s$ , is first pretrained in a standard supervised manner with the labeled source dataset. The employed cross-entropy loss function  $L_{sc}$  is expressed as

$$\min_{\theta_{E_s}} L_{sc}(X_s, X_s) = - \sum_k \mathbb{I}[y_s = k] \log(p_k) \quad (9)$$

$$p_k = \text{Softmax}(f_s^k(x_s)) \quad (10)$$

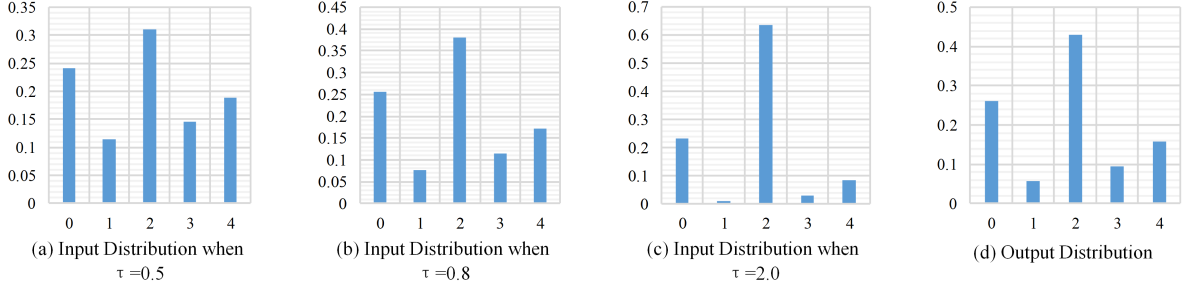


Fig. 6. (a)–(c) Three distributions that are input to the *Softmax* layers with diverse values of  $\tau$  ( $\tau = 0.5, 0.8, 2$ ) when the same probability distribution is output. (d) Output distribution. For better illustration, the three input distributions are shown after *Softmax* ( $\tau = 1$ ). With the same probability distribution output, it can be seen that most intraclass semantic information is retained in the input distribution when  $\tau$  equals 0.5. However, the input distribution in (a) is less discriminative than the distributions in (b) and (c), which is often hard to be classified as a result. Therefore, setting an appropriate value of  $\tau$  can make the classification task easy and retain sufficient intraclass semantic information.

where  $\mathbb{I}[\cdot]$  denotes the one-hot label of the  $k$ th instance;  $f_s^k$  is the source classifier activation of this instance; and  $\theta_{E_s} = [w_{s1}, w_{s2}, \dots, w_{sn}, b_{s1}, b_{s2}, \dots, b_{sn}]$  denotes parameters of  $E_s$ . Then, we initialize the target classification network  $N_t$  (plotted as green blocks in Fig. 1) with the parameters of the optimized  $N_s$ .

2) *Training Target Feature Extractor  $E_t$* : The semantic transfer and the domain adversarial learning are both used to train  $E_t$ . When optimizing  $E_t$ , the parameters of the source feature extractor  $E_s$  are fixed.

Specifically, the parameters of  $E_t$  are optimized with the overall loss  $L_{\text{overall}}$  composed of two parts: supervised adversarial domain loss  $L_{\text{st}}$  and unsupervised semantic transfer loss  $L_{\text{ad}}$ . Hence, the overall loss  $L_{\text{overall}}$  of the *JDS-TL* method can be written as

$$L_{\text{overall}} = L_{\text{ad}} + \alpha L_{\text{st}} \quad (11)$$

where the hyperparameter  $\alpha$  determines how strongly the semantic transfer loss influences the optimization. In this article,  $\alpha$  is set to 0.3 by using greedy search.

Due to the characteristic of the adversarial training mechanism, an additional network  $D$  is trained alternatively with  $E_t$ . As a result, the loss  $L_{\text{overall}}$ , like  $L_{\text{ad}}$ , is composed of two functions, as given by

$$\begin{aligned} \min_{\theta_D} L_{\text{overall}}^D(X_s, X_{ut}, \theta_{E_s}, \theta_{E_t}) \\ = \min_{\theta_D} L_{\text{ad}}^D(X_s, X_{ut}, \theta_{E_s}, \theta_{E_t}); \end{aligned} \quad (12)$$

$$\begin{aligned} \min_{\theta_{E_t}} L_{\text{overall}}^{E_t}(X_s, X_{ut}, X_{lt}, Y_{lt}, \theta_D) \\ = \min_{\theta_{E_t}} \left( L_{\text{ad}}^{E_t}(X_s, X_{ut}, \theta_D) + \alpha L_{\text{st}}(\theta_{E_t}, X_{lt}, Y_{lt}) \right). \end{aligned} \quad (13)$$

By minimizing the two functions alternatively, the parameters of  $D$  and  $E_t$  are optimized, respectively.

During training, the root mean square prop (RMSProp) optimizer is adopted. The learning rate is set to 0.00005. The batch size is set to 32. The pseudocode of the training process of *JDS-TL* is depicted in Algorithm 1.



Fig. 7. T-SNE result of the radar data in “Radar signatures of human activities.” Green marks represent the data of younger persons (age below 50 years old). Red marks represent the data of elder persons (age above 50 years old). The two black marks represent the centroids of the two clusters, respectively.

## IV. IMPLEMENTATION DETAILS

### A. Dataset Description

To evaluate the performance of the proposed *JDS-TL* method, we use the University of Glasgow experimental open dataset “Radar signatures of human activities” [17]. There are six indoor human activities, including drinking from a cup or glass (A1), standing up (A2), bending to pick up an object (A3), sitting down on a chair (A4), falling down (A5), and walking back and forth (A6), which are performed by volunteers whose ages range from 20 to 100. The volunteers younger than 50 perform all six activities. For health considerations, the volunteers who are older than 50 perform all activities apart from “falling down.” All volunteers performed each type of activity thrice.

The data are collected by using an frequency modulated continuous wave (FMCW) radar operating at 5.8 GHz with a bandwidth of 400 MHz and a chirp duration of 1 ms. The collected backscattering signal is a complex time series. The amplitude and phase of the signal are impacted by the electromagnetic characteristics and kinematics of the observed target [18].

### B. Dataset Segmentation

In this dataset, there is limited experimental activity data for old participants. Generally, the young and the elder move

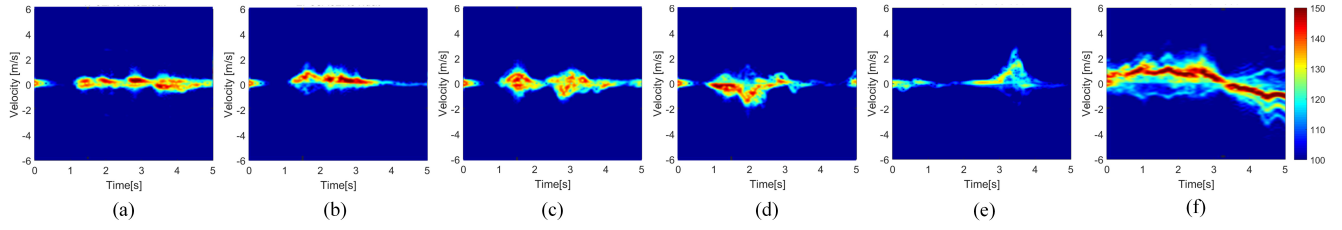


Fig. 8. Several typical radar spectrograms. (a) Drinking from a cup or glass (A1). (b) Standing up (A2). (c) Bending to pick up an object (A3). (d) Sitting down on a chair (A4). (e) Falling down (A5). (f) Walking back and forth (A6).

differently. As a result, there are some differences between the data of the young persons and that of old persons. When a DL model trained with the data of young people is directly used to classify old people's data, the classification performance may drop.

To validate this assumption, we first split the data into two groups, i.e., the data of the volunteers younger than 50 (younger persons) and the data of the volunteers older than 50 (elder persons). Then, we perform t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization on the two groups of data to highlight differences in the data. The t-SNE result is illustrated in Fig. 7. It can be seen that the red cluster and the green cluster have only a little overlap, showing that the data of younger persons and the data of the elder persons have obvious differences. Furthermore, the data of younger persons are not clustered tightly. Some data of younger persons, such as that in the black dotted circle, is away from the centroid of the green cluster.

Then, we choose 70% of the younger persons' data, which is closer to the green cluster centroid, to train a DL classification model. When the trained model is utilized to classify the remaining data, including the data of elder persons and some data of younger persons, its classification performance degrades. This indicates that there are some differences between the data of younger persons and the data of elder persons. Furthermore, there are also some differences within the data of younger persons, due to different experimental environments and motion styles.

Therefore, based on data differences, the dataset is divided into two subsets: the source dataset  $X_s$  and the target dataset  $X_t$ .  $X_s$  consists of 70% of the younger persons' data that is closer to the green cluster centroid in Fig. 7.  $X_t$  consists of the remaining younger persons' data and all the elder persons' data.

Then, to construct a sparsely labeled target dataset, we randomly select 10% instances in  $X_t$  and label them. As for the other instances, the labels are discarded. To get a well-trained DL model for the target domain with limited labeled target instances, we use the source dataset in the proposed *JDS-TL* method for knowledge transfer, as described in Section III.

### C. Radar Data Preprocessing

Radar MD signatures [30] are adopted in this article. An infinite impulse response (IIR) filter is first utilized to remove the static background clutter. Next, the 800-point short time Fourier transform (STFT) with a 0.2 s Hamming window

is performed on the raw radar data to transform the data into a series of 2-D MD spectrograms. Several typical radar spectrograms are shown in Fig. 8. After being resized into  $128 \times 128$  pixels, the spectrograms are input to the proposed *JDS-TL* network. The HAR problem is then solved as an image classification task.

## V. EXPERIMENTAL RESULTS AND DISCUSSION

### A. Classification Results

Since the labeled target instances are selected randomly, the classification performance fluctuates with the variation of the distribution of the selected labeled instances. Thus, we adopt independent repeated experiments to obtain an average accuracy. Specifically, we randomly select 10% target instances from the target dataset  $X_t$  as the labeled subset  $X_{lt}$  for supervised semantic transfer and perform the experiments. The random selection process is repeated 50 times.

1) *Overall Performance of JDS-TL*: With the 50 trials, an average classification accuracy of 87.6% is achieved by using the proposed *JDS-TL*. For comparison, we classify the target instances directly with the source classification model  $N_s$  that is trained with the source dataset (*baseline 1*). The classification results are shown in Fig. 9 with red marks. The average accuracy of recognizing the six activities is approximately 79.4%. Furthermore, when the proposed adversarial domain transfer is applied for DA but the semantic transfer is not used in *JDS-TL* (*baseline 2*), an average accuracy of 84.1% is obtained, as shown in Fig. 9 with yellow marks.

2) *Correlation Between Classification Performance and Distribution of Labeled Target Instances*: To show how the distribution of the labeled target instances affects the classification performance of *JDS-TL*, the correlation between the classification accuracy and the distribution of labeled target instances is shown in Fig. 10. When the normalized variance is between 0.1 and 0.3, *JDS-TL* yields relatively high classification accuracies. This shows that when the labeled data is approximately uniformly distributed, the proposed method can learn sufficient information about every class and obtain good classification performance. When the variance increases, the accuracy declines, showing that uneven data distribution leads to less robust classification performance. However, it is noted that when the normalized variance is between 0.6 and 0.8, the classification exceeded the average again. This is because in this case, though the data distribution is quite uneven, more interclass information can be learned from the



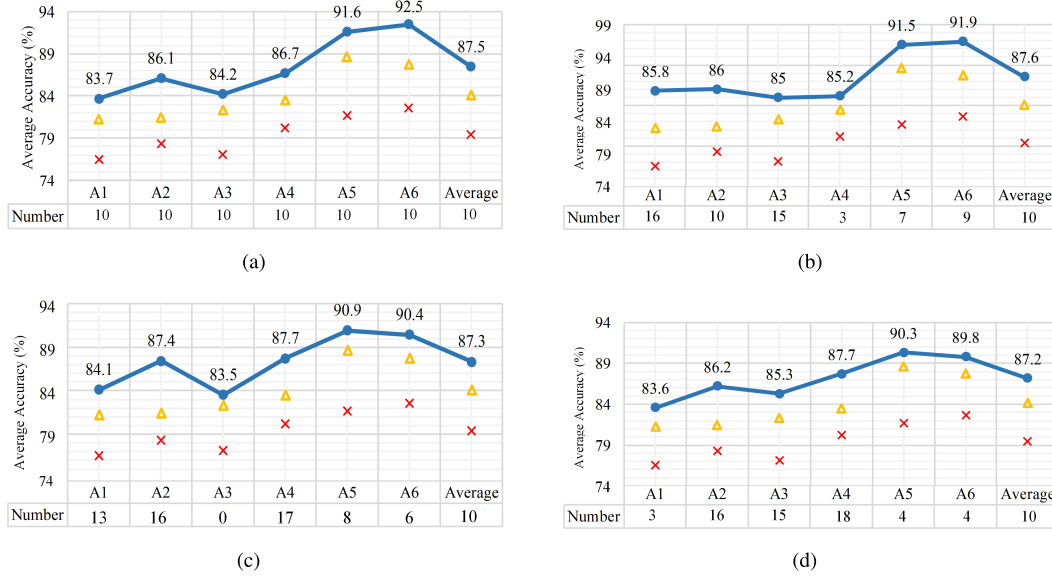


Fig. 9. Classification accuracies of *JDS-TL* with various distributions of labeled target instances. Four typical distributions and the corresponding classification results are shown in the subfigures. The red crosses denote the classification accuracies of directly using  $N_s$  that is trained with  $X_s$  (*baseline 1*). The yellow triangles denote the accuracies of only using the proposed multilayer ADA for DA, and the semantic transfer is not applied (*baseline 2*). (a) Classification result with Distribution 1. (b) Classification result with Distribution 2. (c) Classification result with Distribution 3. (d) Classification result with Distribution 4.

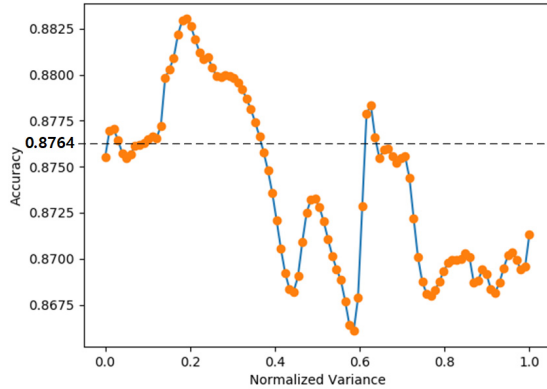


Fig. 10. Correlation between the classification accuracy of *JDS-TL* and the distribution of labeled target instances.

classes that have sufficient samples by using the semantic transfer mechanism.

3) *Typical Cases With Various Target Labeled Data Distributions*: The results of several typical cases with various target data distributions are shown in Fig. 9. It can be seen from Fig. 9 that the proposed *JDS-TL* can bring more improvement on classifying A1, A2, and A3, whose classification results on *baseline 1* are relatively weaker than those of A5 and A6. Specifically, as shown in Fig. 9(a), when the selected instances are approximately evenly distributed, the performance of the proposed method on every type of activity is better than that of *baseline 1* and *baseline 2*. Average accuracy of 87.5% for classifying the six activities is achieved by the proposed method. As shown in Fig. 9(b) and (c), when the distribution is not uniform, the classification performance of *JDS-TL* keeps better than that of the two *baselines*. Average accuracies of 87.6% and 87.3% are achieved under the two distributions (D2 and D3). In particular, for the class with scarce labeled

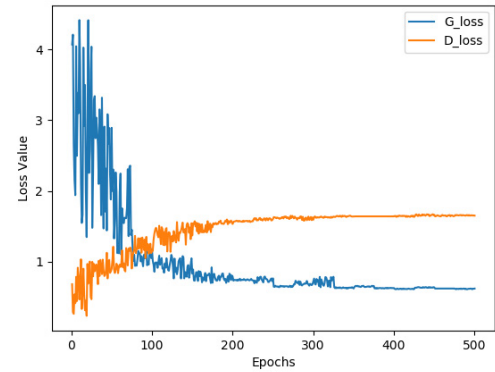


Fig. 11. Loss curves of  $E_t$  and  $D$  when the proposed *JDS-TL* is training. It can be seen that after 100 epochs, the losses of  $E_t$  and  $D$  fluctuate less and tend to stabilize. After approximately 320 epochs, both the losses of  $E_t$  and  $D$  keep stable.

target instances (A4 in D2 and A3 in D3), the classification accuracies do not decrease. As shown in Fig. 9(d), when there are more labeled target instances in A2, A3, and A4, the proposed *JDS-TL* method yields an average accuracy of 87.2%, outperforming *baseline 1* with 7.8% and *baseline 2* with 3.1%. This is because increasing the number of labeled instances can improve classification accuracy. Furthermore, with only 3 instances for A1, an accuracy of 83.6% for classifying A1 is achieved, higher than the accuracies of *baseline 1* and *baseline 2*. This mainly benefits from the extracted interclass information by using the proposed semantic transfer scheme.

4) *Training Loss Curves*: To show the convergence process of the proposed *JDS-TL* during training, the loss curves of the target feature extractor  $E_t$  and the multilayer discriminator  $D$  are illustrated in Fig. 11. It can be seen that after 100 epochs,

TABLE I  
PERFORMANCE COMPARISON IN AVERAGE ACCURACY

No.	Methods	Source data	Target labelled data	Target unlabelled data	Average Accuracy(%)
1	GRL [31]	Yes	No	Yes	81.1 $\pm$ 0.2
2	ADDA [15]	Yes	No	Yes	83.0 $\pm$ 0.1
3	Pretrain+Finetune [32]	Yes	Yes	No	83.3 $\pm$ 3.5
4	SJFT [33]	Yes	Yes	No	83.7 $\pm$ 2.7
5	Pseudo-Labels [34]	Yes	Yes	Yes	82.5 $\pm$ 2.9
6	Manifold Regularization [35]	Yes	Yes	Yes	84.2 $\pm$ 3.0
7	<b>Ours</b>	Yes	Yes	Yes	<b>87.6 <math>\pm</math> 0.3</b>

the losses of  $E_t$  and  $D$  fluctuate less and tend to stabilize. After approximately 320 epochs, both the losses of  $E_t$  and  $D$  keep stable.

### B. Comparison With the State-of-The-Art Methods

To demonstrate the effectiveness of the proposed *JDS-TL* method for HAR with a sparsely labeled dataset, we compare *JDS-TL* with the following state-of-the-art methods.

- 1) *Gradient reversal layer (GRL)* [31] is an unsupervised TL method. A GRL is proposed to alleviate the domain discrepancy between the source and the target domain.
- 2) *Adversarial discriminative domain adaptation (ADDA)* [15] is an unsupervised TL method that utilizes the generative adversarial training scheme for DA.
- 3) *Pretrain+Finetune* [32] is a typical fine-tuning-based supervised TL method. A DL model is first pretrained with the labeled source data, and then the labeled target instances are adopted to fine-tune the pretrained model.
- 4) *Selective joint fine-tuning (SJFT)* [33] is a supervised TL method that uses the labeled source and target data to optimize a DL model. A source data selection mechanism is proposed to select partial source instances to fine-tune the DL model, together with the target data.
- 5) *Pseudo-labels* [34] is a semisupervised method. The pseudo-labels of the unlabeled target data are obtained with a model trained with the source data. Then, the pseudo and labeled datasets are employed to retrain the model together.
- 6) *Manifold regularization* [35] is a semisupervised method that utilizes manifold regularization [36] to train a DL model. Both the labeled and unlabeled target instances are utilized.

We adopt the same 50 target distributions as in Section IV-A to perform the comparison experiments and obtain the average accuracies of the 6 methods under those distributions. The comparison results are shown in Table I. It can be seen that the proposed *JDS-TL* achieves the best performance among these methods, with an average accuracy of 87.6%. In detail, the UDA methods, *GRL* and *ADDA*, obtain average accuracies of 81.1% and 83.0%, respectively. Compared with the unsupervised methods, the supervised TL approaches, *Pretrain+Finetune* and *SJFT*, are more effective for the classification task, with average accuracies of 83.3% and 83.7%, respectively. Regarding the two semisupervised approaches, *pseudo-labels* cannot achieve the expected performance. It is

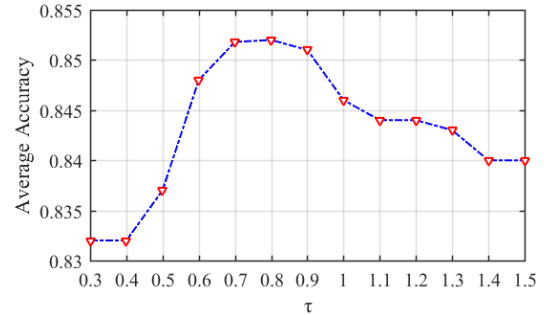


Fig. 12. Variations on average accuracy of *JDS-TL* with diverse values of  $\tau$ . It can be seen that when the value of  $\tau$  is set between 0.7 and 0.8, the proposed *JDS-TL* achieves the best performance, with an accuracy of approximately 85.2%.

due to the fact that the semantic information of the target domain cannot be learned sufficiently when only a small number of labeled target instances are provided. As a result, the assigned pseudo labels differ greatly from the groundtruth.

### C. Ablation Study on *JDS-TL*

1) *Effect of Semantic Transfer*: We first perform the sensitivity analysis of the hyperparameter  $\tau$  in the semantic transfer loss  $L_{st}$  in (8). The performance of *JDS-TL* with diverse values of  $\tau$  is shown in Fig. 12. From Fig. 12, we can see that when the value of  $\tau$  is set between 0.7 and 0.8, the proposed *JDS-TL* achieves the best performance with an average accuracy of approximately 85.2%. When the value of  $\tau$  is lower than 0.7, the average accuracy is more sensitive to the variation of  $\tau$ . When  $\tau$  is larger than 0.8, the average accuracy of *JDS-TL* continues to decline and tends to stabilize at about 84.0%.

We proceed to explore the impact of the number of labeled instances on the soft-label-based semantic transfer. The variation of average accuracies of *JDS-TL* and the two semisupervised methods, *Pseudo-labels* [34] and *Manifold regularization* [35], is illustrated in Fig. 13. Independent repeat experiments are performed under different parameter settings to obtain the average accuracy. For example, in the setting of *Percentage* = 10%, 10% instances are randomly selected from the target dataset  $X_t$  for supervised semantic transfer. The process is repeated 10 times. It is shown that with more labeled target data, the classification accuracies of all three methods increase. *JDS-TL* achieves the highest accuracy when there are no more than 25% target instances labeled, followed by *Manifold regularization*. When there are more

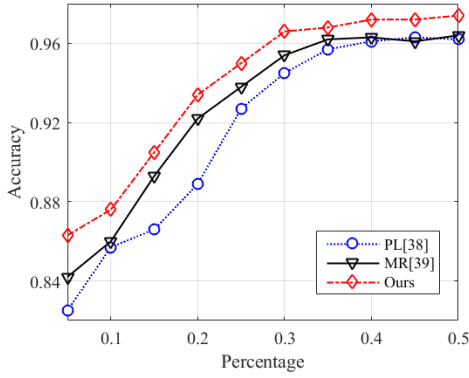


Fig. 13. Variations on average accuracy of *JDS-TL* with the number of labeled target instances.  $x$ -axis represents the percentage of target labeled instances used for supervised semantic transfer, while  $y$ -axis represents the average accuracy of *JDS-TL*.

TABLE II  
PERFORMANCE COMPARISON IN AVERAGE ACCURACY (%)  
OF DIFFERENT ADA METHODS

	w/o DA	w/ DA	Acc. Diff.
GRL [31]	80.4	81.1	+ 0.7
ADDA [15]	81.2	83.0	+ 1.8
<b>Ours</b>	79.4	<b>84.1</b>	<b>+ 4.7</b>

than 30% labeled instances, even if the performance of the three methods is similar, our model consistently yields the best performance. When there are over 50% target instances labeled, the classification accuracy of *JDS-TL* tends to be stable. It is because the proposed method can learn sufficient semantic information for each class from the 50% instances, and new semantic information is extracted less and less.

2) *Effect of Adversarial Domain Transfer*: We evaluate the unsupervised classification performance of the proposed DA method, and compare it with the other two state-of-the-art DA methods, GRL [31] and ADDA [15]. All these methods are configured the same as the references cited, and evaluated on the dataset described in Section IV-A. To remove possible biases, the three methods are repeated 5 times to obtain the average accuracies. Table II shows the average accuracies of the three models under two circumstances: 1) classified with  $N_s$  (w/o DA), 2) classified with  $N_t$  after DA (w/ DA). We see that GRL and ADDA achieve average accuracies of 81.1% and 83.0%, respectively, while our proposed DA method achieves a higher accuracy of 84.1%. By comparing the differences between the accuracy of w/ DA and the accuracy of w/o DA, it is shown that the proposed method provides the best classification performance on the target dataset, with an accuracy difference of 4.7%.

Furthermore, we perform the sensitivity analysis on the hyperparameter  $\gamma$  in (1). The average accuracies of the proposed multilayer DA with diverse values of  $\gamma$  are shown in Fig. 14.<sup>1</sup> It can be seen that when  $\gamma$  is lower than 0.4, the performance of the proposed DA method keeps continually

<sup>1</sup>When  $\gamma$  is set to a quite small value, the model cannot converge. Hence,  $\gamma$  starts at 0.2.

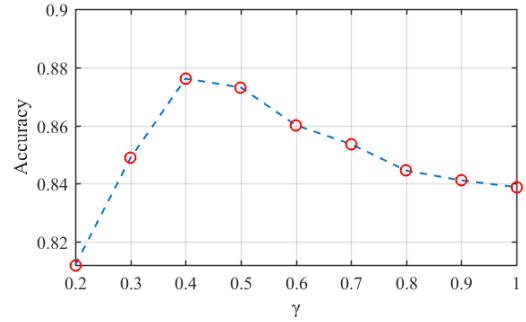


Fig. 14. Variations on average accuracy of *JDS-TL* with diverse values of  $\gamma$ . When  $\gamma$  is lower than 0.4, the performance of the proposed adversarial domain transfer keeps continually improving and achieves the highest average accuracy of approximately 87.6%. When  $\gamma$  is larger than 0.4, the average accuracy decreases.

TABLE III  
AVERAGE ACCURACY OF *JDS-TL* WITH DIFFERENT OF SNR

	0 dB	5 dB	10 dB	15 dB	20 dB	25 dB
Acc.	59.5	70.7	75.5	81.1	87.1	<b>87.6</b>
(%)	$\pm 0.4$	$\pm 0.6$	$\pm 1.3$	$\pm 0.9$	$\pm 0.2$	$\pm 0.3$

improving, and the highest average accuracy of approximately 87.6% is obtained. When  $\gamma$  is larger than 0.4, the average accuracy decreases. As a result, with the heuristic searches, we set  $\gamma$  to 0.4 for experiments.

#### D. Influence of SNR

To evaluate the robustness of the proposed *JDS-TL* method, we add additive white Gaussian noise (AWGN) with diverse signal-to-noise (SNR) levels (0, 5, 10, 15, 20, 25 dB) to the raw radar echo to simulate the noisy MD spectrograms. Then, the proposed *JDS-TL* model is employed to classify those spectrograms. The experiments for each SNR setting are repeated 10 times to obtain an average accuracy. The classification results are listed in Table III. We can see that the proposed method can hardly classify the spectrograms under 0 dB because of the intense noise. When SNR increases to 10 dB, an average accuracy of 75.5% is achieved. Furthermore, the classification performance improves continuously with SNR increasing. When SNR grows to over 20 dB, since the quality of MD spectrograms is good enough, the accuracy increases to over 87.0%.

## VI. CONCLUSION

In this article, we have proposed a new *JDS-TL* method to recognize human activities in a semisupervised manner. The proposed *JDS-TL* is composed of the unsupervised adversarial domain transfer module and the supervised semantic transfer module. The former can mitigate the harmful effects of domain shift and learn domain alignment feature representation. The latter can transfer the intraclass correlation from the source domain to the target domain by utilizing a small amount of labeled target data. A public radar-based HAR dataset including six daily human activities was employed to evaluate the proposed method. Experimental results have proved the efficacy of the *JDS-TL* method, which can achieve an average accuracy of 87.6% when there are merely 10% labeled



instances in the target dataset. Additionally, compared with the fine-tuning-based TL methods, *JDS-TL* is more suitable to classify sparsely labeled data. Further analysis has demonstrated the efficiency of the proposed adversarial domain transfer module and the semantic transfer module. Furthermore, the influence of SNR of radar data on the performance of *JDS-TL* has been explored.

## REFERENCES

- [1] L. Gutiérrez-Madroñal, L. La Blunda, M. F. Wagner, and I. Medina-Bulo, "Test event generation for a fall-detection IoT system," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6642–6651, Aug. 2019.
- [2] S. Stavrotheodoros, N. Kaklanis, K. Votis, and D. Tzovaras, "A smart-home IoT infrastructure for the support of independent living of older adults," in *Proc. IFIP Int. Conf. Artif. Intell. Appl. Innov.*, Rhodes, Greece, May 2018, pp. 238–249.
- [3] Z. Lin, T. Lv, and P. T. Mathiopoulos, "3-D indoor positioning for millimeter-wave massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2472–2486, Jun. 2018.
- [4] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, May 2019.
- [5] C. Ding *et al.*, "Continuous human motion recognition with a dynamic range-Doppler trajectory method based on FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6821–6831, Sep. 2019.
- [6] X. Bai, Y. Hui, L. Wang, and F. Zhou, "Radar-based human gait recognition using dual-channel deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9767–9778, Dec. 2019.
- [7] Y. Yang, C. Hou, Y. Lang, T. Sakamoto, Y. He, and W. Xiang, "Omnidirectional motion classification with monostatic radar system using micro-Doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3574–3587, May 2020.
- [8] Y. He, L. Dai, and H. Zhang, "Multi-branch deep residual learning for clustering and beamforming in user-centric network," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2221–2225, Oct. 2020.
- [9] I. Alnujaim, D. Oh, I. Park, and Y. Kim, "Classification of micro-Doppler signatures measured by Doppler radar through transfer learning," in *Proc. 13th Eur. Conf. Antennas Propag. (EuCAP)*, Krakow, Poland, Apr. 2019, pp. 1–3.
- [10] A. Shrestha *et al.*, "Cross-frequency classification of indoor activities with DNN transfer learning," in *Proc. IEEE Radar Conf. (RadarConf)*, Boston, MA, USA, Apr. 2019, pp. 1–6.
- [11] M. S. Seyfioglu, B. Erol, S. Z. Gurbuz, and M. G. Amin, "DNN transfer learning from diversified micro-Doppler for motion classification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 5, pp. 2164–2180, Oct. 2019.
- [12] H. Du, T. Jin, Y. Song, Y. Dai, and M. Li, "Efficient human activity classification via sparsity-driven transfer learning," *IET Radar, Sonar Navigat.*, vol. 13, no. 10, pp. 1741–1746, Oct. 2019.
- [13] H. Du, T. Jin, Y. Song, and Y. Dai, "Unsupervised adversarial domain adaptation for micro-Doppler based human activity classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 62–66, Jan. 2020.
- [14] Y. Lang, Q. Wang, Y. Yang, C. Hou, D. Huang, and W. Xiang, "Unsupervised domain adaptation for micro-Doppler human motion classification via feature fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 392–396, Mar. 2019.
- [15] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 7167–7176.
- [16] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*. [Online]. Available: <http://arxiv.org/abs/1503.02531>
- [17] D. F. Fioranelli, D. S. A. Shah, H. Li, A. Shrestha, D. S. Yang, and D. J. L. Kerne, "Radar sensing for healthcare," *Electron. Lett.*, vol. 55, no. 19, pp. 1022–1024, Sep. 2019.
- [18] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, Jul. 2019.
- [19] J. L. Kerne *et al.*, "Radar signal processing for sensing in assisted living: The challenges associated with real-time implementation of emerging algorithms," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 29–41, Jul. 2019.
- [20] A. Singh, S. U. Rehman, S. Yongchareon, and P. H. J. Chong, "Sensor technologies for fall detection systems: A review," *IEEE Sensors J.*, vol. 20, no. 13, pp. 6889–6919, Jul. 2020.
- [21] K. Chaccour, R. Darazi, A. H. El Hassani, and E. Andrès, "From fall detection to fall prevention: A generic classification of fall-related systems," *IEEE Sensors J.*, vol. 17, no. 3, pp. 812–822, Feb. 2017.
- [22] Q. Chen, Y. Liu, F. Fioranelli, M. Ritchie, and K. Chetty, "Eliminate aspect angle variations for human activity recognition using unsupervised deep adaptation network," in *Proc. IEEE Radar Conf. (RadarConf)*, Boston, MA, USA, Apr. 2019, pp. 1–6.
- [23] J. Ba and R. Caruana, "Do deep nets really need to be deep?" in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, Sep. 2014, pp. 2654–2662.
- [24] T. Fukuda, M. Suzuki, G. Kurata, S. Thomas, J. Cui, and B. Ramabhadran, "Efficient knowledge distillation from an ensemble of teachers," in *Proc. 18th Annu. Conf. Int. Speech Commun. Assoc.*, Stockholm, Sweden, Aug. 2017, pp. 3697–3701.
- [25] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "FitNets: Hints for thin deep nets," 2014, *arXiv:1412.6550*. [Online]. Available: <http://arxiv.org/abs/1412.6550>
- [26] J. Yim, D. Joo, J. Bae, and J. Kim, "A gift from knowledge distillation: Fast optimization, network minimization and transfer learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 4133–4141.
- [27] C. Yang, L. Xie, S. Qiao, and A. Yuille, "Knowledge distillation in generations: More tolerant teachers educate better students," 2018, *arXiv:1805.05551*. [Online]. Available: <http://arxiv.org/abs/1805.05551>
- [28] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," *IEEE Access*, vol. 7, pp. 36322–36333, 2019.
- [29] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach CA, USA, Jun. 2017, pp. 5767–5777.
- [30] V. C. Chen, "Analysis of radar micro-Doppler with time-frequency transform," in *Proc. 10th IEEE Workshop Stat. Signal Array Process.*, Harrisburg, PA, USA, Aug. 2000, pp. 463–466.
- [31] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, Apr. 2016.
- [32] H. Du, Y. He, and T. Jin, "Transfer learning for human activities classification using micro-Doppler spectrograms," in *Proc. IEEE Int. Conf. Comput. Electromagn. (ICCEM)*, Chengdu, China, Mar. 2018, pp. 1–3.
- [33] W. Ge and Y. Yu, "Borrowing treasures from the wealthy: Deep transfer learning through selective joint fine-tuning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1086–1095.
- [34] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Represent. Learn. (ICML)*, Atlanta, Georgia, vol. 3, Jun. 2013, p. 2.
- [35] R. Gupta, S. Sahu, C. Espy-Wilson, and S. Narayanan, "Semi-supervised and transfer learning approaches for low resource sentiment classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 5109–5113.
- [36] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.



**Xinyu Li** (Student Member, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2017, where she is pursuing the Ph.D. degree with the School of Information and Communication Engineering.

Her main research interests are human detection and activity recognition, machine learning, and small-sample learning.





**Yuan He** (Member, IEEE) received the B.Sc. and M.Sc. degrees from the National University of Defense Technology, Changsha, China, in 2007 and 2010, respectively, and the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2014.

He is an Assistant Professor with the Beijing University of Posts and Telecommunications, Beijing, China. His main research interests are machine learning, signal processing, and electromagnetic computation.



**Xiaojun Jing** (Member, IEEE) received the M.S. and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 1995 and 1999, respectively.

He is a Professor with the Beijing University of Posts and Telecommunications, Beijing, China. His research interests include information security and image processing.



**Francesco Fioranelli** (Senior Member, IEEE) received the Ph.D. degree from Durham University, Durham, U.K., in 2014.

He is Assistant Professor with TU Delft, Delft, The Netherlands, and was an Assistant Professor with the University of Glasgow, Glasgow, U.K., from 2016 to 2019 and the Research Associate with University College London, London, U.K., from 2014 to 2016. His research interests include the development of radar systems and automatic classification for human signatures analysis in healthcare and security, drones

and unmanned aerial vehicles (UAVs) detection and classification, automotive radar, wind farm, and sea clutter.