APSIPA Transactions on Signal and Information Processing, 2024, 13, e301
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, for non-commercial use, provided the original work is properly cited.

Original Paper

Cross-domain Behavior Recognition Based on Millimeter-wave Radar

Rendao Wang[1] and Binquan Wang[2*]

[1] *Institute of Advanced Technology, University of Science and Technology of China*
[2] *School of Cyber Science and Technology, University of Science and Technology of China*

ABSTRACT

Behavior recognition using millimeter wave (mmWave) signals has become a hot topic in recent years. However, existing works are mainly based on the premise that training samples and test samples have the same distribution, which leads to weak robustness of the network model to the environment during actual deployment. In this paper, we propose a domain adaptation framework for action recognition based on mmWave radar signals. Specifically, we use a convolutional neural network to construct our encoder to extract behavioral features in RF signals, use a semi-supervised learning method to pre-train the network, and finally we design a pseudo-label-based fine-grained domain adversarial network to further train the encoder. We conduct extensive experiments on our own collected behavioral data and two publicly available datasets. Experimental results demonstrate the superiority of our method.

Keywords: Human Activity Recognition, Cross Domain, Unsupervised Domain Adaptation

*Corresponding author: Binquan Wang, wbq0556@ustc.edu.cn. This work was supported by National Key R&D Program under Grant 2022YFC0869800 and 2022YFC2503405, National Natural Science Foundation of China under Grant 62201542, 62172381 and 62302471,

Received 31 August 2023; Revised 22 January 2024
ISSN 2048-7703; DOI 10.1561/116.00000262
© 2024 R. Wang and B. Wang

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

1 Introduction

In recent years, with the continuous development of medical healthcare, smart homes, and monitoring technologies, Human Activity Recognition (HAR) has gained increasing attention. Rafferty *et al.* [21] argue that the intensification of population aging and the advancement of whole-house intelligence have heightened the demand for a reliable and secure HAR system. Common HAR technologies currently encompass systems based on wearable devices, such as various methods outlined by S. Zhang *et al.* [32], as well as systems grounded in visual approaches and those reliant on radar sensors, and the various methods summarized by Wu *et al.* [27]. The wearable method necessitates users to don sensor devices, potentially causing discomfort and proving challenging to implement in security and surveillance contexts. Although computer vision methods can achieve high precision in recognizing human activities, as highlighted by X. Wang *et al.* [26], they have also prompted concerns about privacy and security. The radar system can still work normally under conditions such as poor lighting conditions and occluded targets, which shows that it has strong robustness to the environment. And it can well solve the disadvantages of the two methods mentioned above, so using radar as a HAR system is a very promising direction.

Given the remarkable achievements of deep learning technology across various fields, it is evident that within the current context of using millimeter-wave radar for HAR, an increasing number of methods are leveraging deep learning to achieve behavior recognition after processing the raw radar signals. Examples of research contributions in this area include the works of Park *et al.* [20], Z. Chen *et al.* [8], Alkasimi *et al.* [2], and J. Zhu *et al.* [33]. These methods combine the advantages of signal processing and neural network feature extraction to achieve HAR based on radar signals. However, these methods do not consider the issue of domain shift. Since radar echoes contain information about not only human movement but also environmental details, position, angle, and other factors, these specifics comprise the domain information. Thus, when using radar signals for HAR, it is critical to address the challenge posed by domain shift.

Currently, most RF-based domain adaptation methods adopt domain adversarial training methods, such as the methods of B.-B. Zhang *et al.* [29] and Q. Chen *et al.* [7], which only use source domain data for supervised training in the pre-training stage and ignore the target domain. The role of data. In the subsequent adversarial training stage, binary adversarial is used. The domain discriminator only distinguishes whether the input data comes from the source domain or the target domain, ignoring the specific category

Key Research and Development Program of Anhui under Grant 2022h11020026 and the Fundamental Research Funds for the Central Universities.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

of the data. There are also some RF-based domain adaptation methods that use semi-supervised learning methods. For example, the method of B.-B. Zhang *et al.* [30] only uses consistency regularization and minimum entropy regularization to extract unlabeled data features of the target domain. We think this cannot Fully extract domain-independent information. In response to the above two problems, we use the pre-training method of semi-supervised learning to add the target domain data to the pre-training process of the network and mine the representation information of the data in the target domain to better realize the migration and generalization of the model. We use the idea of pseudo-labels to design label values in the domain adversarial training stage, allowing the encoder and domain discriminator to conduct more fine-grained adversarial training and optimizing the effect of adversarial training.

Our main contributions are summarized as follows:

1. We propose a novel cross-domain human activity recognition framework based on millimeter-wave radar signals. Leveraging the concepts of pseudo-labeling and unsupervised domain adaptation, we engineer and embed label values during the domain adversarial phase, followed by employing a fine-grained domain adversarial approach for training.

2. We employ a semi-supervised approach to pre-train the network to address the parameter initialization challenge of fine-grained domain adversarial learning. For pre-training on unlabeled target domain data, we introduce a consistent regularization method.

3. We collect human activity radar data in various environments and encapsulate them into datasets based on the collection scene and location. Extensive experiments are conducted on these datasets and two publicly available radar datasets. Our framework achieves better performance than mainstream domain adaptation methods, demonstrating its superiority.

The rest of this paper is organized as follows. Section 2 introduces the related work. Section 3 formulates the problem, and introduces our domain adaptation training method and related network framework. Section 4 presents the experimental results and analysis, as well as introduces information about the relevant datasets. Section 5 concludes this paper.

2 Related Work

2.1 Cross-Domain Human Activity Recognition Based on Radar Signals

In the process of radar signal propagation, considering factors such as multipath effects and various interferences, the impact of domain shift often becomes

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

more prominent. Currently, a substantial amount of research work has been dedicated to addressing this issue. Lang *et al.* [16] combine manually extracted two types of domain-invariant features with convolutional neural networks to achieve domain adaptation. Q. Chen *et al.* [7] employ adversarial domain training to tackle the domain drift issue caused by different radar placement angles. Khodabakhshandeh *et al.* [15] evaluate the domain adaptation effectiveness of two advanced methods on FMCW radar signals and make some improvements. B.-B. Zhang *et al.* [30] fuse the ideas of consistency regularization loss and pseudo-labels to create a cross-domain model for gesture recognition. Jiang *et al.* [13] construct a domain adaptation framework by combining domain adversarial networks with various constraints and test it on WiFi signals, ultrasound signals, and radar signal data. In this paper, we propose a novel domain training framework. This method first pretrains the network through semi-supervised learning and then, based on the results of domain adversarial training, conducts fine-grained domain adversarial training using the concept of pseudo-labels.

2.2 Semi-Supervised Learning

Semi-supervised learning differs from supervised learning. Based on the perspectives of Van Engelen and Hoos [25], semi-supervised learning involves incorporating unlabeled test set data into the model for training, thereby effectively alleviating the costly process of data annotation. Mainstream semi-supervised learning methods introduce loss terms for unlabeled data, enabling the network to perform well on target domain data. These loss terms can be categorized into three types: entropy minimization, consistency regularization, and generic regularization. Currently, there are several commonly used semi-supervised learning methods, such as MixMatch proposed by Berthelot *et al.* [6] and FixMatch designed by Sohn *et al.* [22]. MixMatch employs the ideas of entropy minimization and consistency regularization, conducting mixed training on labeled and unlabeled data. FixMatch uses pseudo-labeling, where weakly augmented sample outputs are used as pseudo-labels and strongly augmented samples are employed for loss computation. ReMixMatch, studied by Berthelot *et al.* [5], adjusts the label guesses for unsupervised data using the label distribution of supervised data. Simultaneously, it utilizes the predictions from weakly augmented samples as targets for training strongly augmented samples. We plan to use semi-supervised learning methods for network pretraining with the aim of obtaining reliable pseudo-labels for the target domain. This will facilitate subsequent fine-grained domain adversarial training.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

2.3 Unsupervised Domain Adaptation

In the domain of unsupervised domain adaptation based on deep learning, Ben-David *et al.* [4] propose that the source domain consists of labeled samples, while the target domain data comprises unlabeled samples, both contributing to the network training process. Currently, there exist various methods for domain adaptation. Some methods aim to reduce the generalization error in the target domain by minimizing the disparity between the two domains. Examples of such methods include those based on the Maximum Mean Discrepancy (MMD) statistical criterion, such as the Deep Adaptation Network (DAN) introduced by Long *et al.* [18], and the Deep Domain Confusion (DDC) method proposed by Tzeng *et al.* [24], where the latter utilizes multi-kernel MMD. Additionally, certain approaches incorporate the principles of Generative Adversarial Network (GAN), as originally proposed by Goodfellow *et al.* [11], into the framework of unsupervised domain adaptation. Similar to GAN, these methods generally consist of a domain discriminator and a feature extractor. For instance, Adversarial Discriminative Domain Adaptation (ADDA) proposed by Tzeng *et al.* [23] employs separate encoders for the source and target domains. The optimization of the target domain encoder's network parameters is achieved through domain adversarial loss. Domain-Adversarial Training of Neural Network (DANN) introduced by Ganin *et al.* [10] leverages a Gradient Reversal Layer (GRL) to attain the inverse optimization objective for the domain discriminator and the feature extractor. Moreover, certain strategies attempt to make full use of data from both source domain and target domain to directly synthesize target domain samples. These synthesized samples are then used to train the network, thereby achieving domain adaptation. Examples of such methods include CycleGAN proposed by J.-Y. Zhu *et al.* [34], and StyleGAN proposed by Karras *et al.* [14]. In contrast to previous approaches such as ADDA and DANN, our method employs a domain adversarial training strategy. When calculating the domain adversarial loss, we draw inspiration from the approach proposed by Cicek and Soatto [9], which not only considers the domain information of the input data but also incorporates specific class information using a pseudo-labeling technique. We term this approach "fine-grained domain adversarial training."

3 The Proposed Method

This section introduces the proposed domain adaptation framework. The overall network is divided into three parts: feature encoder E, activity recognizer F, and domain discriminator D. The training process comprises two stages. First, E and F are pre-trained using source domain and target domain data through semi-supervised learning. Then unsupervised domain adversarial

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

training is performed based on pseudo-labels to achieve cross-domain behavior recognition based on RF signals.

3.1 Problem Formulation

According to the definition provided by Pan and Q. Yang [19], the input of our model is divided into the source domain D_s and the target domain D_t. In the source domain, the data consists of labeled samples, while in the target domain, the data consists of unlabeled samples. The differences between these domains manifest in the environmental conditions, angles, and locations during data collection. We define the source domain as $(X^s, Y^s) = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$, where y_i^s represents the label corresponding to the ith sample x_i^s, and N_s is the number of samples in the source domain. The set of all unlabeled samples forms the dataset for the target domain, denoted as $X^t = \{x_i^t\}_{i=1}^{N_t}$, where N_t is the number of samples in the target domain. Notably, both the source and target domain data share identical label categories. Our objective is to enhance the network's performance on the target domain by utilizing data from both the source and target domains and employing domain adaptation methods.

3.2 Pre-Training Phase

In the pre-training phase, we optimize the parameters of encoder E and classifier F using source domain and target domain data, as shown in Figure 1. The feature encoder E maps the input data to a feature vector Z, while the classifier F assigns class labels to the features. The goal of the pre-training is to provide relatively accurate pseudo-labels for subsequent domain adversarial training.

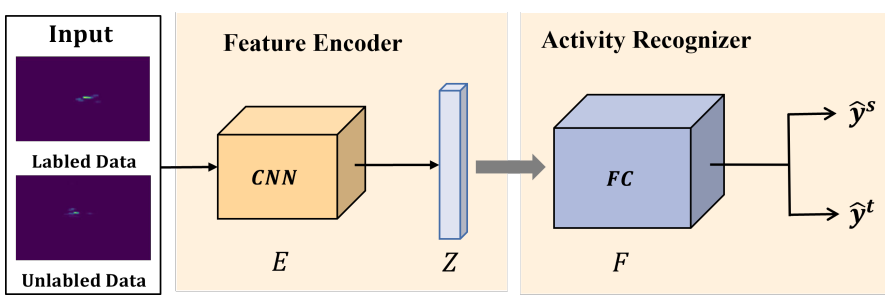

Figure 1: Pre-training model framework.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

3.2.1 Feature Encoder

Convolutional neural networks (CNNs) are effective feature extraction models. In this work, we utilize a CNN as the feature encoder E, as depicted in Figure 2. The encoder comprises three stacked CNN layers. Each layer consists of: a convolution layer with 2D kernels that extracts input features, followed by batch normalization to standardize the data mean and variance, and then max pooling to reduce feature size. At the end, a linear layer maps the output to a high-dimensional vector. We define Θ as the parameter set of the feature extractor. For a given input X, we can obtain its feature representation as:

$$Z = E(X; \Theta). \tag{1}$$

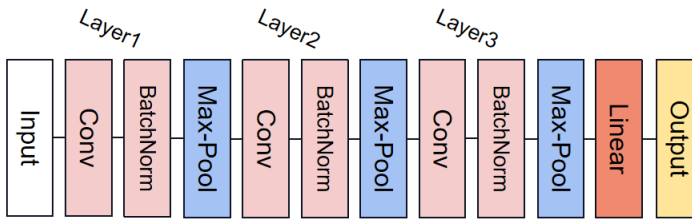

Figure 2: The framework of feature encoder.

3.2.2 Activity Recognizer

To effectively predict labels for human activities, we use a fully connected layer with a nonlinear activation function to map features Z_i to a new latent space $H_i \in \mathbb{R}^C$, where C is the number of categories of labels. And a softmax layer is used to obtain the probability vector of activities as follows:

$$H_i = W_F \cdot Z_i + b_F, \tag{2}$$

$$\hat{y}_i = Softmax(H_i), \tag{3}$$

where W_F and b_F are the parameters of the network, and $\hat{y}_i$ is the predicted result of the network based on the input H_i. Activity Recognizer is essentially a classifier made up of full connections, which we denote as F.

3.2.3 Semi-Supervised Method for Pre-Training

During pre-training phase, we incorporate both source and target domain data. This serves two purposes: first, it allows fully utilizing all available data. Second, since we intend to use a fine-grained domain adversarial approach, we

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

need the encoder E to be able to extract target domain-specific features before conducting this operation. Specifically, we obtain samples from the source domain, denoted as X^s, as well as samples from the target domain, denoted as X^t. After applying data augmentation to the source domain data, we compute the loss function on the source domain data by utilizing the cross-entropy function, which involves the network's output values and their corresponding labels. The loss function on the source domain data is expressed as:

$$L_s = -\frac{1}{|X^s|}\sum_{i=1}^{|X^s|}\sum_{c=1}^{C} y_{ic}^s log\left(\hat{y}_{ic}^s\right), \tag{4}$$

where $|X^s|$ represents the number of labeled samples, y_i^s and $\hat{y_i^s}$ denote the true label and the network's predicted category for the input x_i^s.

For unlabeled samples X^t, according to the consistency regularization method proposed by Abuduweili *et al.* [1], we believe that the network should produce the same output distribution for samples before data augmentation and samples after data augmentation. By computing the L2 distance between the network's output values before and after data augmentation, we derive the loss function on the target domain data:

$$L_t = \frac{1}{|X^t|}\sum_{i=1}^{|X^t|}\sum_{c=1}^{C}\left(y_{ic}^t - \hat{y}_{ic}^t\right)^2, \tag{5}$$

where $|X^t|$ represents the number of unlabeled samples, and y_{ic}^t and $\hat{y}_{ic}^t$ denote the predictions of the network before and after data augmentation for the input x_{ic}^t, respectively. Ultimately, the optimization objective during the pre-training stage is to combine L^s and L^t to yield:

$$L_\alpha = L_s + \alpha L_t, \tag{6}$$

where α is the weight parameter of L^t, and L_α represents the optimization objective for networks E and F.

During the training process, we employ the sharpening method proposed by Berthelot *et al.* [6] to encourage the model to generate high-confidence predictions on unlabeled data, thereby achieving entropy minimization. Additionally, we utilize the MixUp technique introduced by H. Zhang *et al.* [31] to perform joint training by combining data from the source domain and the target domain.

3.3 Domain Adversarial Training Phase

In this stage, we primarily refine the network parameters of our encoder E by employing adversarial training between the domain discriminator D and

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

the encoder E. This process aims to endow E with domain adaptability. The framework of our approach is depicted in Figure 3. Conventional domain adversarial methods commonly use binary adversarial loss to determine if features are from the source or target domain. In contrast, our method introduces an adversarial loss involving $2C$ categories within the network. To achieve this idea, we design the label for the domain adversarial stage, where the initial C categories pertain to the source domain, and the subsequent C categories belong to the target domain, as illustrated in Figure 4. This enables the adversarial training to not only help the encoder learn domain differences, but also differentiate between similar actions across domains, achieving a fine-grained adversarial effect.

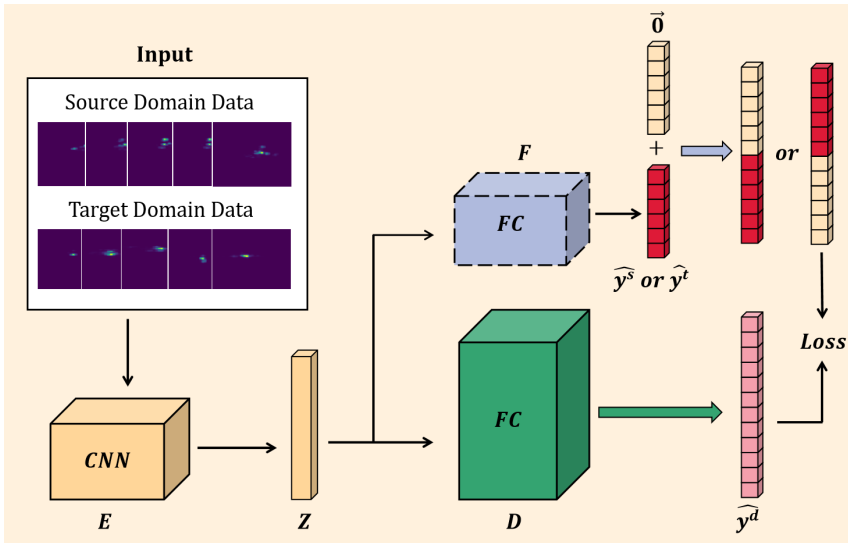

Figure 3: Fine-grained domain adversarial training framework.

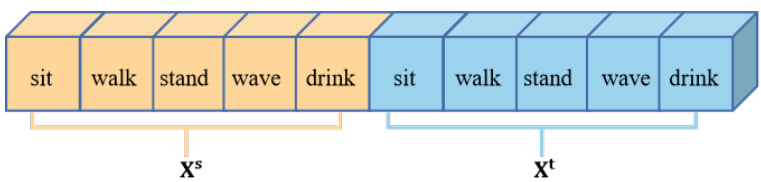

Figure 4: Concatenation of labels from source and target domain aata.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

3.3.1 Domain Discriminator

The output value dimension of activity recognizer F is equal to the total number of categories C:

$$\hat{y}_i = F\left(E\left(x_i\right)\right) \in \mathbb{R}^C. \tag{7}$$

The output value of the domain discriminator D is twice the number of categories $2C$:

$$\hat{y_i^d} = D\left(E\left(x_i\right)\right) \in \mathbb{R}^{2C}. \tag{8}$$

Similar to the concept of classifier F, D gives the judgment of the behavior category of the feature vector. The first half is the possibility of the category on the source domain, and the second half is the possibility of the category of the target domain, as shown in the Figure 4.

3.3.2 Fine-grained Domain Adversarial Training

Throughout the entire process of domain adversarial training, we implement an unsupervised training approach, utilizing both source domain data and target domain data to update networks E and D. Through the preceding pre-training, our encoder E has already developed a preliminary ability to extract valuable features on the target domain. Additionally, classifier F can provide reasonably accurate predictions for samples in the target domain. We utilize these predictions as pseudo-labels during the fine-grained domain adversarial training phase. By concatenating these pseudo-labels with zeros, we construct $2C$-dimensional labels for domain adversarial training. Our optimization objective is derived by computing the loss against the output values of the domain adversarial network D. In the conventional domain adversarial method, the domain discriminator will distinguish the input features into the source domain and the target domain, which is a two-classification problem. The domain discriminator and the feature encoder are trained against each other so that the feature encoder can extract domain-independent features. The fine-grained domain adversarial network transforms the original two-class classification into a $2C$ category classification problem, where C represents the specific number of categories, the first C bits represent the category of the source domain, and the last C bits represent the category of the target domain. We first use the pseudo-label method to encode the labels for adversarial training based on the pre-training results, as shown in Figure 4. For the input data, the domain discriminator will give its predicted value, including which domain it comes from and the specific behavioral category, so a fine-grained effect can be achieved when adversarial training with the feature encoder.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

To mitigate potential instability from adversarial training, we adopt several strategies proposed by Arjovsky *et al.* [3]. Specifically, we use Mean Squared Error (MSE) for the loss function. We also replace the Adam optimizer with stochastic gradient descent (SGD). Furthermore, we remove the sigmoid activation from the last layer of discriminator D.

For updating network D in our methodology, the domain discriminator should accurately determine if inputs are from the source or target domain, and also distinguish between specific behavior categories. First, we concatenate the pseudo-label values used to train D on the source data. The loss on the source domain data using MSE is:

$$L_{ds}(D) = \frac{1}{|X^s|} \sum_{i=1}^{|X^s|} \left(D\left(E\left(x_i^s\right)\right) - \left[\hat{y_i^s}, \vec{0}\right] \right)^2, \tag{9}$$

where the $\vec{0}$ represents a C-dimensional vector, and $\hat{y_i^s}$ is the network's predicted value for the input x_i^s. We concatenate these to form pseudo-labels for the domain discriminator, resulting in $\left[\hat{y_i^s}, \vec{0}\right]$. For the target domain data, given the input x_i^t, we can concatenate to form its pseudo-label from the pre-training stage, which becomes $\left[\vec{0}, \hat{y_i^t}\right]$. We calculate the loss function on the target domain as follows:

$$L_{dt}(D) = \frac{1}{|X^t|} \sum_{i=1}^{|X^t|} \left(D\left(E\left(x_i^t\right)\right) - \left[\vec{0}, \hat{y_i^t}\right] \right)^2. \tag{10}$$

In domain adversarial training, the primary role of the feature encoder is to extract domain-invariant features, thereby preventing the domain discriminator from distinguishing the categories of these features. For the source domain data, the pseudo label is $\left[\vec{0}, \hat{y_i^s}\right]$, and for the target domain data, the pseudo label is $\left[\hat{y_i^t}, \vec{0}\right]$. Consequently, when updating the feature encoder E, we can formulate the loss functions for the source domain, L_{es}, and for the target domain, L_{et}, as follows:

$$L_{es}(E) = \frac{1}{|X^s|} \sum_{i=1}^{|X^s|} \left(D\left(E\left(x_i^s\right)\right) - \left[\vec{0}, \hat{y_i^s}\right] \right)^2, \tag{11}$$

$$L_{et}(E) = \frac{1}{|X^t|} \sum_{i=1}^{|X^t|} \left(D\left(E\left(x_i^t\right)\right) - \left[\hat{y_i^t}, \vec{0}\right] \right)^2. \tag{12}$$

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

Finally, we can get the overall loss function, where Formula 13 represents the loss function of the domain discriminator, and Formula 14 represents the loss function of the feature encoder. During the domain adversarial phase, the parameters of Networks D and E are updated in an alternating manner.

$$L_d(D) = L_{ds} + L_{dt}, \tag{13}$$

$$L_e(E) = L_{es} + L_{et}. \tag{14}$$

4 Experiment

In this section, we first collect human activity data in various environments using FMCW radar. Experiments are then conducted on this dataset and two publicly available mmWave radar datasets to evaluate the performance of the proposed network. Finally, ablation experiments are performed to study the effects of pre-training and fine-grained domain adversarial techniques.

4.1 Baseline Methods

We compare our method to three different deep domain adaptation models: DANN, ADDA and MixMatch. DANN and ADDA employ domain adversarial strategies, while MixMatch uses a semi-supervised approach. These three methods are highly regarded as classic domain adaptation frameworks and widely applied in radar-based behavior recognition, with ideas referenced by Jiang *et al.* [13], Khodabakhshandeh *et al.* [15], and B.-B. Zhang *et al.* [30]. For comprehensive comparison, we use identical feature encoders and classifiers, and tailor distinct domain discriminators to handle the varying domain information in each method.

4.2 Data Asugmentation

In order to make better use of existing data and improve the generalization ability of the training model, data enhancement is very necessary. We adopt a consistent conventional semi-supervised learning method in the pre-training stage, and in order to reduce the difference between the original features and the enhanced features, we perform data augmentation by randomly adding noise points. After preprocessing the original radar signal, for the features on each timestamp, we randomly select some feature points and set them to zero:

$$\mathbb{F}^t_{(i,j)} = 0, t = 1, 2, \ldots T, (i, j) \in Rand(m) \tag{15}$$

where $\mathbb{F}^t_{(i,j)}$ is the feature of t-th time dimension, Rand(m) denotes local feature coordinates to be erased and m is the number of erasing feature.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

4.3 Experiment with our Dsataset

4.3.1 Dataset

We invite 7 volunteers to collect radar data of their various actions at different locations. The data collection process involves capturing six common actions: standing, walking, squatting, bending, waving, and drinking water. Figure 5 illustrates distinct indoor environments comprising laboratory setups, office spaces, and corridor scenarios. Each environment varies in dimensions and furnishing arrangements, presenting distinct domain-specific information. Data collection also encompasses diverse angles within the same environment. For the laboratory setting, radar placements are altered concerning subjects' positions and angles, we select three locations centered on the subject: $(1.8m, 0°)$, $(1.8m, 30°)$, and $(1.8m, 60°)$. In indoor settings, two distinct angles and positions are chosen: $(1.8m, 0°)$ and $(1.8m, 30°)$. In corridor scenarios, the radar is positioned directly in front of the subjects. Consequently, six distinct environment configurations are established for the datasets. Each volunteer is instructed to perform each specific action roughly 30 times at each location. The entire data collection spans 20 days, yielding a cumulative total of 11,805 data samples. To facilitate subsequent descriptions, we define symbolic representations for datasets with distinct domain information, as outlined in Table 1.

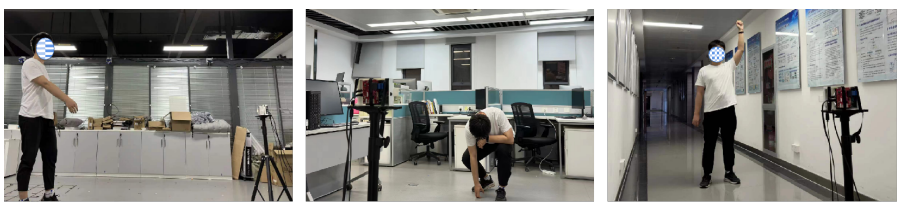

Figure 5: The environment of data collection, from left to right is the laboratory, office and corridor.

4.3.2 Device Configuration and Signal Processing

We utilize the TI AWR1843 mmWave radar along with the DCA1000 real-time data acquisition board to gather radar data of various human behaviors. Our radar signal collection equipment is shown in Figure 6. The left side of Figure 6 is the front side of the circuit board, we can see the radar array with 3 transmitters and 4 receivers, to the right is the back of the board.

Considering the characteristics of the TI AWR1843 mmWave radar, we process the raw radar signals into Dynamic Range Angle Images (DRAI). We employ a 3D-FFT algorithm on the raw signals to extract range, speed, and

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

Table 1: Experimental configurations on our dataset.

Symbolic Name	Domain Information
Dataset A	Laboratory Location $(1.8m, 0°)$
Dataset B	Laboratory Location $(1.8m, 30°)$
Dataset C	Laboratory Location $(1.8m, 60°)$
Dataset D	Office Location $(1.8m, 0°)$
Dataset E	Office Location $(1.8m, 30°)$
Dataset F	Corridor Location $(1.8m, 0°)$

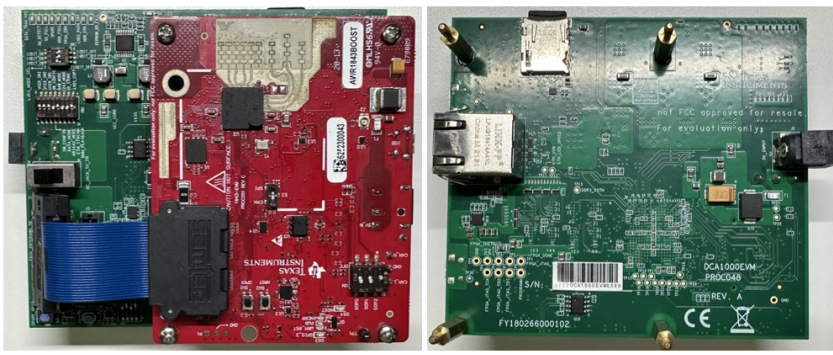

Figure 6: Radar signal acquisition board.

Figure 7: The calculation process of DRAI sequences.

angle information of human activities, as illustrated in Figure 7. Subsequently, in the Doppler dimension, we suppress static clutter by zeroing out information below a velocity threshold. Due to real-time constraints and the complexity of subsequent network inputs, we perform a weighted sum of this 3D tensor along the Doppler dimension to reduce data size. Despite the loss of speed information, we stack 40 frames of Range Angle Images (RAI) together to form a DRAI. This DRAI serves as the input to the encoder E.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

4.3.3 Results and Discussion

We evaluate the cross-domain capability of the proposed framework, including factors such as different environments, perspectives, etc. For each set of experiments, all unlabeled target domain data are used for training, and all target domain data are used for testing. We will use all the baseline methods to compare with our method, the relevant experimental configuration and experimental results are shown in Table 1. The meaning of AB → C is: dataset A and dataset B are used as source domain data, and dataset C is used as target domain data.

The configuration and results of each group of experiments are shown in the Table 2, and the results of baseline method and the proposed framework are compared. We will analyze the model's cross-perspective and cross-environment capabilities based on experimental results. The first three sets of experiments in Table 2 mainly evaluate the cross-angle capability between different models. We take the radar data collected from two angles as the source domain, and the other is the target domain data. All three experiments show that our method is more accurate. Among them, data set A and data set C are collected from the front and side of the collected subjects respectively, and they have better domain adaptability for data set B collected in the oblique direction, and our method achieves an accuracy rate of 92.82% on this. It is worth noting that we processed the raw radar data into DRAI, which is highly correlated with the relative distance and angle between the human body and the radar. And angle B is between angle A and angle C. We have reason to believe that the data of angle A and angle C contain information about angle B, and the positive human behavior data contains high-quality behavioral feature information, ultimately resulting in higher experimental results for AC *to*B than the other two.

The later sets of experiments in the table show the cross-environment capabilities of different models. We take the data of the laboratory environment as the source domain, combine the data of the office environment and the channel environment in different combinations, and construct the source domain data and the target domain data. The results also show that the framework domain proposed by us has stronger adaptability. In summary, we have done extensive experiments on this dataset to verify the ability of our model across angles, locations, and environments, and our method has the highest classification accuracy compared to other methods with an average accuracy of 66.89%, demonstrating the superiority of our framework.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

Table 2: Domain adaptation experiments and results (%).

Methods	DANN	ADDA	MixMatch	Ours
AB→C	56.11	54.35	55.22	**58.77**
AC→B	76.92	74.43	91.53	**92.82**
BC→A	57.87	46.54	61.95	**63.15**
AD→F	45.78	46.23	48.96	**50.14**
ABC→F	52.27	52.38	46.36	**53.95**
ABC→DE	70.86	69.21	70.96	**73.54**
ABCDE→F	54.03	52.59	65.75	**67.34**
ABCF→DE	71.01	70.43	72.64	**75.42**
Avg	60.61	59.52	64.17	**66.89**

4.4 Experiment with the Gesture Recognition Dataset

4.4.1 Dataset

Gesture actions, distinct from activities like walking and bending, involve smaller-scale movements performed by the human body. To evaluate our framework, we employ the gesture recognition dataset based on millimeter-wave radar, which was publicly released by Li *et al.* [17]. This dataset is collected from 25 volunteers positioned across six distinct environments and five different locations. Figure 8 showcases various indoor environments and depicts the radar placement within each environment. The dataset features the selection of six common gestures and also includes the collection of other human behaviors as negative samples. Each volunteer is instructed to perform any type of gesture 5 or 10 times at each location, resulting in a total of 10,650 gesture samples and 13,400 negative samples. In this experiment, we employ distinct rooms as criteria for categorizing data in different fields, with the corresponding symbols defined in Table 3.

The radar equipment uses for this data collection is the same as ours, which is the AWR1843 radar. So for data preprocessing, we use the same process of generating DRAI. Considering the characteristics of gestures relative to normal human activities, in subsequent data processing we trim the feature map size and slightly reduce the time window.

4.4.2 Results and Discussion

In this round of experiments, we evaluate our framework on publicly available gesture recognition dataset. Since the dataset are mainly collected in different

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

Figure 8: The environment in which the data is collected.

Table 3: Experimental configurations on the gesture recognition dataset.

Symbolic Name	Domain Information
Dataset A	Room 1
Dataset B	Room 2
Dataset C	Room 3
Dataset D	Room 4
Dataset E	Room 5
Dataset F	Room 6

environments, we design several experiments to test the cross-environment capabilities of our framework, mainly by configuring the ratio of source domain data to target domain data. The details and symbol definitions for this dataset are shown in Table 3. During the experiment, the data from both source domain and target domain participates in the training of the network. After the training, the tagged target domain data is used to participate in the evaluation of the network. The experimental results are shown in Table 4. Compared with other methods, our framework achieves the highest average accuracy of 97.25 %, indicating that this approach can significantly improve model performance in the target domain.

4.5 Experiment with the Open Human Activity Radar Dataset

4.5.1 Dsataset

In this experimental section, we utilize a publicly available dataset provided by Guendel *et al.* [12]. This dataset is based on radar data collected from five

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

Table 4: Domain adaptation experiments and results (%).

Methods	DANN	ADDA	MixMatch	Ours
ABCDE→F	92.53	89.87	92.12	**94.45**
ABCDF→E	95.75	96.57	96.82	**97.99**
ABCD→EF	94.08	90.52	96.29	**97.63**
ABEF→CD	95.17	94.76	97.03	**97.82**
ABC→DEF	94.17	93.39	96.25	**97.37**
DEF→ABC	94.77	93.21	97.09	**98.21**
Avg	94.41	93.05	95.93	**97.25**

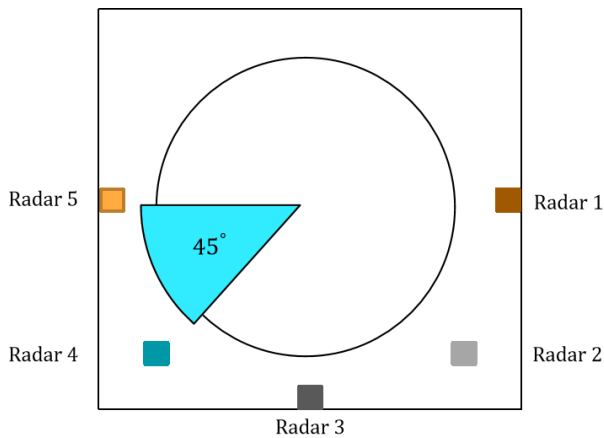

Figure 9: Radar placement of open dataset.

distinct positions within an indoor environment. Figure 9 illustrates these five radar placement positions along with their varying orientations and angles. The dataset is gathered from 15 different volunteers and covers a range of 10 distinct activities, including walking, standing, and sitting. As the tenth activity class in the dataset is categorized as chaotic, we collect a total of 54,720 samples from the well-defined first nine activity classes for our experiment. Notably, these activities are performed as continuous sequences of actions. In this experiment, we utilize the radar's angular position as an indicator to differentiate data from different domains. The symbolic definitions of the dataset and their corresponding domain information are presented in Table 5.

In this dataset, a single transmitter and a single receiver are placed at each position, and the collected radar data can obtain the target's distance, time and Doppler information in three fields. Range is the measured distance of the

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

Table 5: Experimental configurations on the open dataset.

Symbolic Name	Domain Information
Dataset A	Radar 1
Dataset B	Radar 2
Dataset C	Radar 3
Dataset D	Radar 4
Dataset E	Radar 5

target from the radar, and Doppler represents the radial velocity of the target. We can get the distance between the target and the radar by calculating the delay from transmission to reception. In a pulsed UWB radar, each received radar echo is digitized to generate fast-time samples representing the distance of targets. By performing measurements over multiple pulse repetitions in the slow-time domain, it becomes possible to generate a Range Time Map (RTM) for each Pulse Repetition Interval (PRI). X. Yang *et al.* [28] utilize this dataset and processed it into RTMs to accomplish the task of behavior classification. In this experiment, we also use RTM to implement behavior recognition.

4.5.2 Results and Discussion

We use the data preprocessing method announced by the producer of the dataset to process the raw data into an RTM heatmap, and then we convert it to an RGB image and use the image as the input to the network. This round of experiments mainly evaluates the cross-angle ability of our model. We design multiple sets of experiments, similar to the previous configuration, we designate one of the angles as the target domain, the rest of the angles as the source domain data, and experiments with different ratios of source and target data. The dataset configuration and results are shown in Table 6. The average accuracy of domain adaptation of our framework at a single angle is 82.65%, and the average accuracy on all domain adaptation experiments is 83.79%, both of which are the highest values among all methods. The experiments demonstrate the superior performance of our method. Different from the dataset we collected, the direction of human movement in the public dataset used in this section is random, causing the domain adaptability of the model in Section 4.3 to be greatly affected by the radar placement angle. This also leads to differences in the performance of the model on the two datasets.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

Table 6: Domain adaptation experiments and results (%).

Methods	DANN	ADDA	MixMatch	Ours
ABCD→E	76.22	79.53	80.06	**82.31**
ABCE→D	75.21	77.84	79.67	**81.53**
ABDE→C	80.43	82.48	83.78	**84.11**
ABC→DE	77.74	76.56	78.71	**80.27**
ABD→CE	82.83	82.93	86.08	**87.37**
ABE→CD	81.93	79.46	85.73	**87.16**
Avg	79.06	79.80	82.34	**83.79**

4.6 Ablation Experiment

4.6.1 The impact of the pre-training process and domain adversarial training process on the model

We propose a framework that includes semi-supervised pre-training and unsupervised domain adversarial training. To investigate the contributions of these two training stages to overall performance, we plan to conduct ablation experiments on our collected radar dataset and the gesture recognition dataset. We refer to the network trained solely through direct fine-grained domain adversarial training as Bare Model 1, and the network trained only through semi-supervised pre-training as Bare Model 2. We compare the results of domain adaptation from these two models with those of the fully trained Full Model. The experimental setup and results are presented in Table 7 and Table 8, where Table 7 presents ablation results on our dataset, while Table 8 presents results on the gesture recognition dataset.

Table 7: Results of ablation experiments on our dataset (%).

Methods	Bare Model 1	Bare Model 2	Full Model
ABC→F	20.76	47.34	**53.95**
ABC→DE	22.65	71.07	**73.54**
ABCDE→F	24.43	65.98	**67.34**
ABCF→DE	21.84	72.65	**75.42**

By comparing the results of Bare Model 1 and Full Model, it becomes evident that the network trained directly through fine-grained domain adversarial training performs significantly poorer. This is attributed to the unsupervised nature of our domain adversarial training, where the encoder and classifier

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

Table 8: Results of ablation experiments on the gesture recognition dataset (%).

Methods	Bare Model 1	Bare Model 2	Full Model
ABCDE→F	48.22	92.12	**93.45**
ABCDF→E	47.61	96.82	**97.99**
ABCD→EF	35.09	96.29	**96.63**
ABEF→CD	38.83	97.03	**97.82**
ABC→DEF	23.06	96.25	**97.37**
DEF→ABC	33.45	97.09	**98.21**

initially lack the capacity to effectively extract features and classify accurately. The absence of precise pseudo-label support leads to unstable training and hindered convergence. On the other hand, comparing Bare Model 2 with Full Model, we observe a substantial improvement in performance for the Full Model.

Subsequent fine-grained domain adversarial training will further refine the outcomes of the pre-trained network, thereby achieving the highest overall cross-domain performance. The semi-supervised pre-training approach effectively initializes our network, providing a strong foundation for subsequent fine-grained domain adversarial training.

4.6.2 Impact of target domain data size

Considering that the target domain data scale may have an impact on model performance, we conducted another ablation experiment to test the impact of different target domain data scales on performance. Similar to the above experiment, we conduct ablation experiments on our own dataset and a public gesture recognition dataset. We mainly conduct the ABC→DE experimental group on our data set, and the ABC→DEF experimental group on the gesture recognition data set, adjusting the scale of the target domain to 100%, 80%, 70%, 50% and 25%, the experimental results are shown in Table 9. From the experimental results, we can see that as the amount of data in the target domain is reduced, the performance of the model will not change significantly. Too little target domain data will have an impact on the performance of the model.

4.7 Qualitative Asnalysis Esxperiment

In this section, we first conduct comparative experiments using the experimental group ABC→DE in Table 2 to compare the cross-domain capabilities of

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

Table 9: Results of ablation experiments (%).

Dataset size	100%	80%	70%	50%	25%
Our dataset	73.54	71.40	69.27	70.79	66.33
Gesture recognition dataset	97.37	97.19	96.51	95.24	93.08

the proposed model and the cross-domain capabilities after fine-tuning. The result after fine-tuning the network parameters is 81.43%, the result after domain adaptation is 73.54%, and the result of using only labeled training is 62.39%. Considering that we cannot use target domain data for fine-tuning in practical scenarios, the model we proposed has excellent domain adaptability and application prospects.

The we use the encoder E trained by the experiment ABCF→DE in Table 7 to extract features from the target domain data, and draw t-SNE graphs based on these feature vectors, as shown in Figure 10. The left picture shows the result without using the domain adaptation method, and the right picture shows the result with the domain adaptation method. Similarly, we draw the t-SNE graph on the gesture recognition dataset based on the training results of DEF→ABC in Table 8 as shown in Figure 11. It is not difficult to see that through our domain adaptation method, the feature encoder E can effectively extract the behavioral information of the target domain. Compared with the results in the left figure on the right figure, the feature vectors of each category are clustered closer to each other, and the distance between the feature vectors of different categories becomes larger, which effectively proves that our method has excellent domain adaptability.

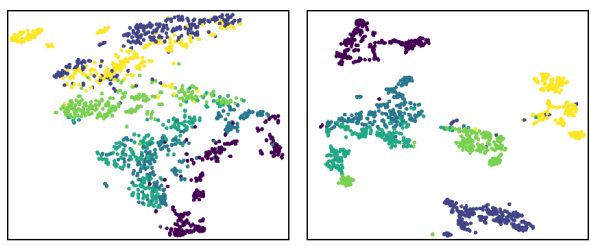

Figure 10: T-SNE graphs before and after domain adaptation (using our dataset).

5 Conculsion

In this paper, we propose a more general unsupervised domain adaptation framework for human activity recognition using radar signals. Our framework

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

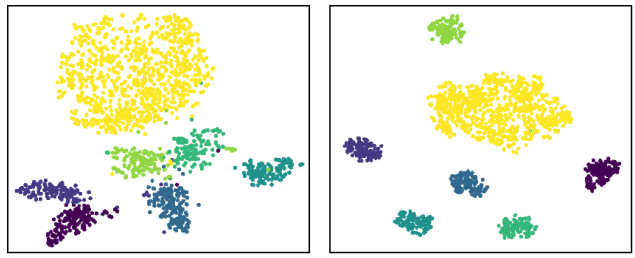

Figure 11: T-SNE graphs before and after domain adaptation (using the gesture recognition dataset).

consists of an encoder, a classifier and a domain discriminator. The network is pre-trained through semi-supervised learning. The purpose is to obtain more accurate pseudo-labels in the target domain, and then perform fine-grained unsupervised domain confrontation to complete the overall frame training. We collect radar behavior data in different environments to build a dataset, and conduct experiments on this dataset and two publicly available radar datasets to evaluate the performance of our framework. Experimental results demonstrate the effectiveness and superiority of the proposed framework.

References

[1] A. Abuduweili, X. Li, H. Shi, C.-Z. Xu, and D. Dou, "Adaptive consistency regularization for semi-supervised transfer learning", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, 6923–32.

[2] A. Alkasimi, A.-V. Pham, C. Gardner, and B. Funsten, "Human Activity Recognition Based on 4-Domain Radar Deep Transfer Learning", in *2023 IEEE Radar Conference (RadarConf23)*, IEEE, 2023, 1–6.

[3] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks", in *International conference on machine learning*, PMLR, 2017, 214–23.

[4] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation", *Advances in neural information processing systems*, 19, 2006.

[5] D. Berthelot, N. Carlini, E. D. Cubuk, A. Kurakin, K. Sohn, H. Zhang, and C. Raffel, "Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring", *arXiv preprint arXiv:1911.09785*, 2019.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

[6] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "Mixmatch: A holistic approach to semi-supervised learning", *Advances in neural information processing systems*, 32, 2019.

[7] 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 *2019 IEEE Radar Conference (RadarConf)*, IEEE, 2019, 1–6.

[8] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks", *IEEE Geoscience and Remote Sensing Letters*, 15(5), 2018, 669–73.

[9] S. Cicek and S. Soatto, "Unsupervised domain adaptation via regularized conditional alignment", in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, 1416–25.

[10] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks", *The journal of machine learning research*, 17(1), 2016, 2096–30.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks", *Communications of the ACM*, 63(11), 2020, 139–44.

[12] R. G. Guendel, M. Unterhorst, F. Fioranelli, and A. Yarovoy, "Dataset of continuous human activities performed in arbitrary directions collected with a distributed radar network of five nodes", *4TU. ResearchData*, 10(16691500), 2021, v3.

[13] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, *et al.*, "Towards environment independent device free human activity recognition", in *Proceedings of the 24th annual international conference on mobile computing and networking*, 2018, 289–304.

[14] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, 4401–10.

[15] H. Khodabakhshandeh, T. Visentin, R. Hernangómez, and M. Pütz, "Domain Adaptation Across Configurations of FMCW Radar for Deep Learning Based Human Activity Classification", in *2021 21st International Radar Symposium (IRS)*, IEEE, 2021, 1–10.

[16] 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 Geoscience and Remote Sensing Letters*, 16(3), 2018, 392–6.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

[17] Y. Li, D. Zhang, J. Chen, J. Wan, D. Zhang, Y. Hu, Q. Sun, and Y. Chen, "Towards domain-independent and real-time gesture recognition using mmwave signal", *IEEE Transactions on Mobile Computing*, 2022.

[18] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks", in *International conference on machine learning*, PMLR, 2015, 97–105.

[19] S. J. Pan and Q. Yang, "A survey on transfer learning", *IEEE Transactions on knowledge and data engineering*, 22(10), 2009, 1345–59.

[20] J. Park, R. J. Javier, T. Moon, and Y. Kim, "Micro-Doppler based classification of human aquatic activities via transfer learning of convolutional neural networks", *Sensors*, 16(12), 2016, 1990.

[21] J. Rafferty, C. D. Nugent, J. Liu, and L. Chen, "From activity recognition to intention recognition for assisted living within smart homes", *IEEE Transactions on Human-Machine Systems*, 47(3), 2017, 368–79.

[22] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence", *Advances in neural information processing systems*, 33, 2020, 596–608.

[23] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, 7167–76.

[24] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance", *arXiv preprint arXiv:1412.3474*, 2014.

[25] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning", *Machine learning*, 109(2), 2020, 373–440.

[26] X. Wang, Z. Xia, H. Wang, and F. Xu, "Human Behavior Recognition Based on Multi-Dimensional Feature Learning of Millimeter-Wave Radar", in *2021 Signal Processing Symposium (SPSympo)*, IEEE, 2021, 284–8.

[27] D. Wu, N. Sharma, and M. Blumenstein, "Recent advances in video-based human action recognition using deep learning: A review", in *2017 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2017, 2865–72.

[28] X. Yang, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Radar-based human activities classification with complex-valued neural networks", in *2022 IEEE Radar Conference (RadarConf22)*, IEEE, 2022, 1–6.

[29] B.-B. Zhang, D. Zhang, Y. Li, Y. Hu, and Y. Chen, "Unsupervised domain adaptation for device-free gesture recognition", *arXiv preprint arXiv:2111.10602*, 2021.

[30] B.-B. Zhang, D. Zhang, Y. Li, Y. Hu, and Y. Chen, "Unsupervised domain adaptation for device-free gesture recognition", *arXiv preprint arXiv:2111.10602*, 2021.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026

[31] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization", *arXiv preprint arXiv:1710.09412*, 2017.

[32] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, and N. Alshurafa, "Deep learning in human activity recognition with wearable sensors: A review on advances", *Sensors*, 22(4), 2022, 1476.

[33] J. Zhu, H. Chen, and W. Ye, "A hybrid CNN–LSTM network for the classification of human activities based on micro-Doppler radar", *Ieee Access*, 8, 2020, 24713–20.

[34] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks", in *Proceedings of the IEEE international conference on computer vision*, 2017, 2223–32.

Downloaded from http://ftp.nowpublishers.com/atsip/article-pdf/13/4/1/11379406/116.00000262en.pdf by guest on 02 July 2026