This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3256324

Mean Teacher-based Cross-Domain Activity Recognition using WiFi Signals

Chunjing Xiao, Yue Lei, Chun Liu and Jie Wu, *Fellow, IEEE*

***Abstract*—WiFi Channel State Information (CSI)-based activity recognition has initiated a great many studies because of wide availability and privacy protection. However, general recognition approaches still struggle to generalize beyond the source domain of training data, i.e., well-trained models might not be suitable to target data with unseen subjects or environments. Existing solutions, such as few-shot learning-based and data augmentation-based approaches, either require a few labeled target samples which is difficult to be collected especially for old target users, or inappropriately treat augmented samples with different amounts of noise. To overcome these limitations, we propose a Mean Teacher-based cross-domain human activity recognition framework using WiFi CSI, WiTeacher. In this framework, to address shift between source and target domains, we built a label smoothing-based classification loss, where the input data are the target-like samples generated by StyleGAN, and corresponding label values are dynamically adjusted by our designed adaptive label smoothing method. To enhance the model robustness, we devise a sample relation-based consistency regularization term to keep the distances of the two samples with and without perturbations invariant, which can exploit the relationships between samples to improve recognition performance. The experiments illustrate that WiTeacher achieves obvious gains without requiring any annotation data from the target domain.**

***Index Terms*—Human activity recognition, WiFi channel state information, Mean Teacher, label smoothing.**

I. Introduction

Human activity recognition systems are key elements of many emerging Internet of Things applications, such as smart buildings, identification, health care, etc [1], [2]. Recently, many works have made a substantial process in this field. Among them, WiFi Channel State Information (CSI)-based activity recognition is regarded as a promising technique because it can achieve non-restrictive, privacy-friendly activity sensing when compared to current technologies such as cameras or wearable sensors [3], [4]. These privileges are attracting a lot of works to present different solutions for various applications.

The typical CSI-based activity recognition models are subject-specific, i.e., they might not perform well for target users whose data are not adopted for model training [5], [6], [7], [8]. The reason behind this is that CSI variations induced by different persons exhibit individual differences because of various body characteristics and behavior habits. The model, that is trained without using target data, cannot capture characteristics of target subjects, and correspondingly may suffer from performance degradation. However, practical scenarios are often confronted with this situation that labeled data of target (terminal) users cannot be adopted for model training since it is hard to let terminal users to report the exact activity boundaries and activity categories.

Currently, there are mainly two ways to address this issue: few-shot learning-based [9], [10], [11] and data augmentation-based [12], [13], [14] approaches. The former first pre-trains a model based on the data from the source domain whose labeled samples are easy to be collected, and then fine-tunes the pre-trained model based on a few labeled samples from the target domain to make the model applicable to the target data. And the data augmentation-based methods typically introduce Generative Adversarial Networks (GANs) to generate a large number of target-like samples, and then adopt these generated samples to enhance generalization capacity of the classifier to the target data. Both kinds of methods have obtained remarkable improvement on cross domain activity recognition.

Despite of significant success of these methods, there still exist some shortages. First, few-shot learning-based methods still need a few labeled samples from the terminal users [15], [16]. But collecting a few labeled samples is still difficult, especially for old terminal users. Second, data augmentation-based methods generally consider all the generated data to possess the same quality. However, GANs are typically unstable and prone to failure [17], [18], and correspondingly generated samples may exhibit various levels of quality, i.e., some may be like real samples and others may be quite noised. Third, existing methods only consider each sample separately during model training, and ignore the relationships between samples, which can be explored to enhance model robustness.

To address these issues, we present a Mean Teacher-based cross-domain activity recognition framework using WiFi CSI, WiTeacher. In this framework, we introduce Mean Teacher as the basic semi-supervised classification framework, and design a new label smoothing-based classification loss and a sample relation-based consistency regularization term to boost activity recognition on cross-domain scenarios. Here, we regard data from a selected user as the target domain and other data as the source domain under a given environment. Specifically, we exploit Mean Teacher to implement activity recognition without requiring any labeled data of target users. Mean-

Manuscript received January 10, 2022; revised December 13, 2022; accepted February 28, 2023. This work is supported in part by the National Natural Science Foundation of China under Grant 62072077, and in part by the Science and Technology Foundation of Henan Province of China under Grant 212102210387. *(Corresponding author: Chun Liu.)*

Chunjing Xiao, Yue Lei and Chun Liu are with the School of Computer and Information Engineering, Henan University, and also with the Henan Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng 475004, China (e-mail: chunjingxiao@gmail.com, leiyue971109@gmail.com, liuchun@henu.edu.cn).

Jie Wu is with Department of Computer and Information Sciences, Temple University, 1805 N. Broad Street, Philadelphia, PA 19122 (e-mail: jiewu@temple.edu).

teacher is an effective way to utilize unlabeled data to enhance the model's generalization capacity by progressively training a detection model in a student-teacher framework. And it has shown supplementary superiority for semi-supervised learning in image processing fields, such as object detection [19], [20] and person re-identification [21], [22].

In cross-domain scenarios, the Mean Teacher model can be easily biased towards the source domain [20] as the supervision is mainly based on the source data. Therefore, we introduce StyleGAN [23] to generate target-like samples based on the labeled source data, which are further assigned certain labels to conduct supervised training to relieve the bias. Although the target-like samples can preserve the main identify information of the original source samples [20], directly assigning the original (hard) labels of the source data to the target-like data is inappropriate, because the generated target-like samples can be noised or distorted due to unstable nature of GANs [17], [24]. Hence, we design an adaptive label smoothing method to dynamically adjust label values according to the quality of generated samples. The adjusted labels, called soft labels, can more accurately reflect category information of generated data. As a result, the target-like data with soft labels are used to build a classification loss to reduce the deviation induced by improper hard labels during supervised training, and further enhance recognition results.

In general Mean Teacher models, each sample is treated separately during model training, i.e., the model only consider the consistency of each single data point, and ignore the relationship between samples. This might lead to the result that the model is smooth in the vicinity of each data point but not smooth in the vacancy among data points [25]. Therefore, we design a sample relation-based consistency regularization term for CSI data. The main idea is to force the distance of the two samples to be consistent with the one of their corresponding augmented ones. This regularization term can make the model more robust and further improve recognition performance.

We summarize the main contributions of this paper as follows:

- We present a Mean Teacher-based cross-domain activity recognition framework using WiFi CSI, WiTeacher, which can effectively conduct cross domain activity recognition without requiring any annotation data from the target domain.
- We design an adaptive label smoothing method to dynamically adjust the label values for the target-like samples generated by StyleGAN. And these target-like samples with adjusted labels are adopted to build a classification loss for enhancing the model's generalization capacity.
- We propose a consistency regularization term about sample relation to keep the distances of the two samples with and without perturbations invariant, which can enhance model robustness.
- Experiment results based on the two public datasets illustrate that our framework outperforms the state-of-the-art approaches in cross domain scenarios. The data and code are available online[1].

[1]https://github.com/ChunjingXiao/WiTeacher

II. Preliminaries

In this section, we present a preliminary overview of Mean Teacher and Label Smoothing. And these will serve as background or key design ingredients of our WiTeacher framework.

A. Mean Teacher

Mean Teacher [26] is initially proposed for semi-supervised learning. It consists of two models with identical architecture, a student model and a teacher model. The student model is trained using the labeled data as standard, and the teacher model uses the exponential moving average weights of the student model. Each sample prediction of the teacher model can be seen as an ensemble of the student model's current and earlier versions, therefore it is more robust and stable. By enforcing the consistency of teacher and student models using a consistency loss based on unlabeled samples, the student model is then guided to yield better performance.

Formally, in the standard setting of semi-supervised learning, we have access to labeled dataset from the source domain $X_L = \{(x_l^i, y^i)\}_{i=1}^M$ and unlabeled dataset from the target domain $X_u = \{x_u^i\}_{i=1}^N$. For labeled samples, the supervised cross entropy (classification) loss is defined as:

$$\mathcal{L}_{lab} = -\frac{1}{M}\sum_{i=1}^{M} y^i \log f_s(x_l^i) \tag{1}$$

where M is the number of the labeled samples and $f_s(x_l)$ refers to the prediction of the student network for the input x_l. For unlabeled samples, the consistency loss penalizes the difference between the prediction $f_s(x_u)$ of the student network and $f_t(x_u)$ of the teacher network. Correspondingly, the loss of unlabeled data is typically computed as Mean Squared Error:

$$\mathcal{L}_{unlab} = \frac{1}{N}\sum_{i=1}^{N} \| \left(f_s(x_u^i) - f_t(x_u^i)\right) \|^2 \tag{2}$$

where N is the number of the unlabeled samples. The total training loss in Mean Teacher is composed of the supervised loss $\mathcal{L}_{lab}$ and consistency loss $\mathcal{L}_{unlab}$.

B. Label Smoothing

Label smoothing regularization [27] is first proposed to improve the generalization and learning efficiency of a neural network by replacing the one-hot vector labels with the smoothed labels that average the hard targets and the uniform distribution of other labels. Specifically, for a K-class classification problem, the one-hot label is smoothed by

$$y^{LS} = (1-\theta)y + \theta y' \tag{3}$$

where y is the one-hot label, $\theta \in (0, 1)$ is the smoothing strength and $y' = 1/K$ is a uniform distribution for all labels. And then the cross-entropy loss will adopt this smoothed label, rather than the original hard one-shot label. This can help to overcome the overfilling issue and improve generalization capacity of the model. And its effectiveness has been demonstrated in many state-of-the-art models for image classification [28], [29], natural language process [30], [31] and vehicle and person re-identification [32], [33].

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3256324

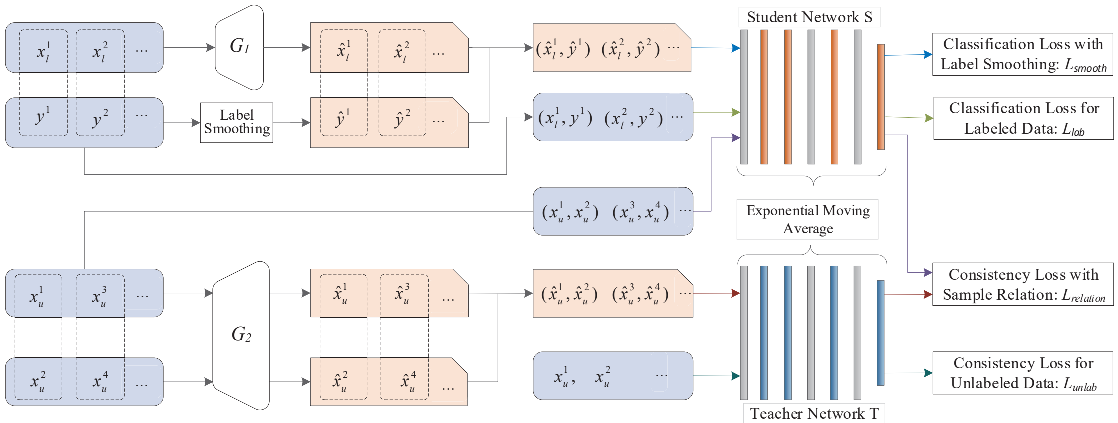


Fig. 1: WiTeacher Framework. Here (x_l, y), ($\hat{x}_l$, $\hat{y}$), x_u and $\hat{x}_u$ refer to source samples with labels, generated target-like samples with soft labels, unlabeled target samples and generated source-like samples respectively. During the training process, first, x_l is transferred into target-like sample $\hat{x}_l$ by generator G_1 and y into soft label $\hat{y}$ by our designed adaptive label smoothing method. And ($\hat{x}_l$, $\hat{y}$) and (x_l, y) are adopted to build the label smoothing-based classification loss $\mathcal{L}_{smooth}$ and standard classification loss $\mathcal{L}_{lab}$ individually. Second, x_u is transferred into source-like sample $\hat{x}_u$ by generator G_2, and sample pairs (x_u^i, x_u^{i+j}) and ($\hat{x}_u^i$, $\hat{x}_u^{i+j}$) extracted from them are used to build the sample relation-based consistency loss $\mathcal{L}_{relation}$. Besides, x_u is utilized to construct the standard consistency loss $\mathcal{L}_{unlab}$.

III. The Proposed Method

In this section, we propose a Mean Teacher-based cross-domain activity recognition method using WiFi CSI, WiTeacher. First, we present an overview of the WiTeacher framework. And next, we describe how to boost the Mean Teacher model via label smoothing based on generated target-like samples and via sample relation based on unlabeled data. Finally, we illustrate the overall training process.

A. Overview of the proposed framework

We present the WiTeacher framework to apply the model to the target data without requiring labeled target samples. The proposed framework is shown in Figure 1, which is mainly composed of the four parts: (1) the label smoothing-based classification loss $\mathcal{L}_{smooth}$ in Equation 8, (2) the standard classification loss $\mathcal{L}_{lab}$ in Equation 1, (3) the sample relation-based consistency loss $\mathcal{L}_{relation}$ in Equation 9, and (4) the standard consistency loss $\mathcal{L}_{unlab}$ in Equation 2. $\mathcal{L}_{lab}$ and $\mathcal{L}_{unlab}$ are derived from the standard Mean Teacher model, and $\mathcal{L}_{smooth}$ and $\mathcal{L}_{relation}$ are our designed losses for enhancing recognition performance on target data.

To alleviate the bias towards the source domain in the standard Mean Teacher framework, we design a label smoothing-based classification loss $\mathcal{L}_{smooth}$. Specifically, we adopt StyleGAN [23] to produce target-like data $\hat{x}_l$ based on the source data x_l. As the hard label y of the source data x_l is inappropriate for the generated data $\hat{x}_l$ which might contain noises or distortions, we design an adaptive label smoothing method to transfer hard label y into soft label $\hat{y}$. And then the soft label is assigned to the generated target-like sample to form labeled target-like sample $(\hat{x}_l, \hat{y})$, which is used to conduct supervised training with classification loss $\mathcal{L}_{smooth}$ to reduce the bias.

To enhance model robustness, we further devise a sample relation-based consistency loss $\mathcal{L}_{relation}$. Unlike existing works that treat each sample separately in the consistency regularization, this loss forces the distance of two samples computed via the student network to keep consistent with the one via the teacher network. In particular, we randomly extract sample pair (x_u^i, x_u^{i+1}) from unlabeled target samples and transfer them into source-like sample pair $(\hat{x}_u^i, \hat{x}_u^{i+1})$ by generator G_2. And then these sample pairs are utilized to construct loss $\mathcal{L}_{relation}$. By introducing the label smoothing-based classification loss and the sample relation-based consistency loss, numerous target data can be adopted for model training, which enables WiTeacher to capture characteristics of target data and extract distinguishable features for activity classification.

B. Booting Mean Teacher with Label Smoothing

Due to the difficulty of collecting labeled data from the target domain, the supervised training for the Mean Teacher model is mainly based on the labeled data from the source domain. And the model can be easily biased towards the source domain and might not appropriately deal with the target data [20]. To alleviate this bias, we first translate the source samples with annotations into target-like ones using StyleGAN [23]. And then we propose an adaptive label smoothing method to transfer hard labels of source data into soft labels for generated target-like data. Compared with hard labels, soft labels can more accurately represent category information of target-like data which might contain noises and distortions because of unstable nature of GANs [17], [24]. Finally, these target-like data with soft labels are adopted to conduct supervised training on the model to reduce the bias towards the source data.

Specifically, we introduce StyleGAN [23] to generate target-like samples based on source samples with annotations. StyleGAN is a generative adversarial network based on style transfer [23], which adds a number of additional modules to the generator to control the style of the generated samples at different levels of detail. By adjusting the hyper-parameters, we adopt the source data and unlabeled target data to train a StyleGAN model with the function of style translation. And then this model is used to transfer the source sample into the target-like one. Since the target-like sample is derived from the source one, a similar label of the source sample can be assigned to the target-like one. However, the hard label of the source sample cannot accurately reflect category information of the generated sample, because the translation process of StyleGAN for the cross-domain style might cause a difficulty in completely preserving identify information and produce noises or distortions in the translated samples [17], [24].

To exactly represent category information, inspired by label smoothing and soft target techniques [27], [34] we propose an adaptive label smoothing method to dynamically adjust the label value according to the quality of the generated sample. And the sample quality is determined by the quantity of noises in the generated target-like samples. When the generated samples contain more noises, its label distribution should be more smoother, and vice versa. Below we first state how to compute the score of sample quality, and then build the classification loss using the generated samples with new soft labels.

Formally, for sample x, assume that its one-hot label is $y = (y_{(1)}, y_{(2)}, ..., y_{(K)})$, meaning that $y_{(i)}$ is 1 for the correct class and 0 for the rest. To accurately reflect the category information of samples, we try to assign a proper value determined by the sample quality to $y_{(i)}$ instead of only using 0 or 1. For a generated sample, we compute its quality score based on two values: one is the similarity degree between this sample and the source samples with the same category; and another is the similarity degree between this sample and the associated target samples. The higher the similary degree is, the bigger the quality score is. The reason is that the generated sample with high quality should have few noises and approach the real one, and further should be quite similar to the real source samples and target samples with the same category. Here, Dynamic Time Wrapping (DTW) is used to compute the similarity degree, for the DTW method is remarkably effective in analyzing the similarity of time series data using a warping path that can detect identical shapes at different time steps [2], [3].

For generated sample $\hat{x}^i$ which is derived from x^i, the DTW distance between this sample and the corresponding source samples is computed as:

$$D^i_{source} = \frac{1}{|X^i_l|} \sum_{t \in X^i_l} DTW(\hat{x}^i, t) \tag{4}$$

where X^i_l refers to the set of the source samples which have the same category with x^i, and $DTW(\hat{x}^i, t)$ refers to the DTW distance between $\hat{x}^i$ and t. Here we adopt the average distance of the samples in X^i_l in order to make computation results of the distance more robust. For the DTW distance between $\hat{x}^i$ and the target samples, because there are no labels for the target samples and $\hat{x}^i$ should be more similar to the ones with the same category, we select the top Z similar samples with $\hat{x}^i$ as the ones with the same category. Hence, the distance between generated sample $\hat{x}^i$ and the corresponding target samples is defined as:

$$D^i_{target} = \frac{1}{Z} \sum_{t \in X^i_u} DTW(\hat{x}^i, t) \tag{5}$$

where X^i_u refers to the set of the top Z similar samples with $\hat{x}^i$. As a result, the quality scores of generated sample $\hat{x}^i$ is calculated by combining these two distances:

$$D^i_{quality} = \beta / D^i_{source} + (1 - \beta) / D^i_{target} \tag{6}$$

where β is the tradeoff parameter to balance the two terms. The closer to a real one the generated sample is, the smaller the DTW distance between them is and the higher the quality score is.

Further, the smoothed soft label for the generated sample $\hat{x}^i$ can be computed based on the quality score as follows:

$$\hat{y} = \overline{D}^i_{quality} \cdot y + (1 - \overline{D}^i_{quality}) \cdot y' \tag{7}$$

where $\overline{D}^i_{quality}$ denotes the normalized $D^i_{quality}$ with the range of [0,1], y is the one-hot label for its corresponding source sample and $y' = 1/K$ is a uniform distribution for all labels.

The proposed label distribution $\hat{y}$ is further used as the soft label of the generated sample for model training. Correspondingly, the classification loss using the generated target-like sample with the soft label becomes:

$$\mathcal{L}_{smooth} = -\frac{1}{|\hat{X}_u|} \sum_{i=1}^{|\hat{X}_u|} \hat{y}^i \log f_s(\hat{x}^i) \tag{8}$$

where $\hat{X}_u$ is the set of the generated samples, and the classifier f_s is the student function that maps input feature space to the label space. In this way, this loss can alleviate the bias towards the source domain and help the model to capture the characteristics of target data, and further improve recognition performance.

C. Booting Mean Teacher with Sample Relation

In the typical mean teacher framework, each unlabeled sample is fed into both the student network and the teacher network, which is trained by minimizing the distance outputted by both networks, such as the works [20], [35]. However, these methods only consider the perturbations around each single sample, while ignore the connections between samples. To exploit relation information, we design a consistency regularization term about sample relation to enhance model robustness. The main idea is that the distance between two samples with perturbations should keep consistent with the distance of the original ones. And the designed regularization term tries to force both distances invariant. Considering the relationship of samples can be regarded as a kind of data augmentation, which can enhance the generalization capacity of the model.

Specifically, we adopt two kinds of data for this regularization term: unlabeled data from the target domain and generated data by StyleGAN. Using unlabeled target data for this regularization term can help the model to learn characteristics

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3256324

of target data, and further advocate generalization capacity of the model. The generated samples are produced by a trained StyleGAN model which takes unlabeled target samples as inputs and generates source-like samples. These source-like samples contains the features of both source and target data, which can enhance data diversity and further improve recognition performance.

Formally, let (x_u^i, x_u^{i+1}) be the real sample pair selected from unlabeled target data, $(\hat{x}_u^i, \hat{x}_u^{i+1})$ be the generated sample pair selected from data produced by StyleGAN, and $h_s(.)$ and $h_t(.)$ be the output vectors of the student network and the teacher network, individually. We adopt the cosine similarity degree to measure the sample relation. Assume that $S(h^i, h^{i+1})$ denotes the cosine similarity degree of the two vectors. Then, the consistency regularization term for the sample relation $\mathcal{L}_{relation}$ is defined as:

$$\begin{aligned}&\mathcal{L}_{relation} = \\ &\frac{1}{F}\sum_{i=1}^{F} \|S\left(h_s(x_u^i), h_s(x_u^{i+1})\right) - S\left(h_t(x_u^i), h_t(x_u^{i+1})\right)\| \\ &+ \|S\left(h_s(\hat{x}_u^i), h_s(\hat{x}_u^{i+1})\right) - S\left(h_t(\hat{x}_u^i), h_t(\hat{x}_u^{i+1})\right)\|\end{aligned} \tag{9}$$

Where F refers to the number of selected sample pairs. For a pair of samples, this term forces the similarity degree of both samples computed by the student network to be close to the one by the teacher network. By considering the correlation between samples, this term can improve the model robustness and further improve recognition performance.

D. Overall Model

To enhance classifier robustness, our proposed label smoothing-based classification loss in Equation 8 and sample relation-based consistency loss in Equation 9 are incorporated into the Mean Teacher model. Coupled with the classification loss in Equation 1 and consistency loss in Equation 2, the final loss function of the model is illustrated as:

$$L = \mathcal{L}_{lab} + \lambda_1\mathcal{L}_{unlab} + \lambda_2\mathcal{L}_{smooth} + \lambda_3\mathcal{L}_{relation} \tag{10}$$

where λ_1, λ_2 and λ_3 are parameters for adjusting the weights of different losses.

Before training WiTeacher, we first generate the source-like and target-like data using StyleGAN offline. For this purpose, we train two generators using the labeled source data and unlabeled target data by leveraging hyper-parameters in StyleGAN. One is used to produce target-like data based on the source data. And another is adopted to generate source-like data based on the target data. Next, the soft labels of the generated target-like samples are computed offline based on the labeled source data and unlabeled target data. Finally, taking these generated data and real data as inputs, the model with the loss in Equation 10 is optimized by the Adam algorithm [36].

IV. Experimental Evaluation

In this section, we evaluate the effectiveness of our proposed WiTeacher compared to the state-of-the-art baselines on two datasets. Also, we exploit the ablation study and the role of training data size. Finally, we investigate the parameter sensitivity on both datasets.

A. Experiment Data and Setup

1) Datasets: We conduct the experiments on two CSI-based behavior recognition datasets. **SignFi data** [37] consists of thousands of CSI traces about sign language gestures, which are frequently used in daily life. And each of the 4 users makes the sign gestures with each gesture repeated for 10 times. **DeepSeg data** [38] is composed of thousands of human activities from 5 volunteers with different body shape and age. Each volunteer is asked to perform every activity thirty times. For all the following experiments, we conduct cross-domain validation by extracting data of different users in a given environment as test data. In other words, we select one person's data as the test data, and others as the training data. To meet the requirement of the semi-supervised model, a number of unlabeled samples from the target domain are selected for model training. Note that these unlabeled data should cover all the categories of the source data.

2) Baselines: To illustrate the effectiveness of our designed framework, we compare WiTeacher with the shallow learning-based semi-supervised methods (e.g., S3VM [39] and Semi-RF [40]), GAN-based methods (e.g., ManiGAN [41] and CsiGAN [13]), Mean Teacher-based methods (e.g., Mean-Teacher [26] and LCMTeacher [42]), and meta learning-based method (e.g., MetaActivity [9]).

- **S3VM** [39]: A widely used non-deep learning-based semi-supervised method. S3VM cannot take the amplitude of CSI data as input, and thus we extract features using the approach in [43] to feed this model.
- **Semi-RF** [40]: A semi-supervised learning algorithm that puts a self-training wrapper on the random forest classifier. Semi-RF takes the same input as the S3VM model.
- **ManiGAN** [41]: A GAN-based semi-supervised learning method incorporating manifold regularization. The method exhibits obvious merits on image classification compared to other GAN-based and non-GAN-based semi-supervised methods (e.g., Local GAN [44] and Virtual Adversarial Training [45]).
- **CsiGAN** [13]: A GAN-based activity recognition model using WiFi CSI. This method introduces a new complement generator and optimizes the outputs and loss functions of the discriminator to improve performance of activity recognition.
- **MeanTeacher** [26]: A Mean Teacher-based semi-supervised learning framework. This model maintains a teacher model's weights as the exponential moving average of a student model's weights and minimizes the divergence between their probability predictions under diverse perturbations of the inputs.
- **LCMTeacher** [42]: An effective Mean Teacher-based framework considering local consistency assumption. This model introduces a local clustering method to alleviate the confirmation bias issue in the Mean Teacher framework, and yields significant improvements compared to Mean Teacher.

- **MetaActivity** [9]: A meta learning-based adaptable activity recognition model. This approach is specifically designed for recognizing activities across scenes and categories using WiFi CSI.

3) Experimental Settings: We adopt the average accuracy and F1-score as the metric for evaluation. Here the accuracy refers to the percentage of activities whose class labels are correctly classified. The F1-score combines Recall and Precision with an equal weight. The model is optimized by Adam with learning rate 0.05, and the mini-batch size of data is 120. The hyper-parameters β and F are empirically set 0.5 and 15%, individually, for both datasets. And Z is set 5 and 15 for SignFi and DeepSeg data respectively.

B. Overall Performance

We first illustrate the effectiveness of our proposed WiTeacher compared to the seven baselines. The results of the two datasets are presented in Figure 2 and 3. From these results, we have following observations.

The first observation is that the deep learning-based methods obviously outperform the two shallow learning-based approaches. Meanwhile, among these deep learning-based methods, the two cross domain recognition methods, WiTeacher and MetaActivity, acquire gains for both datasets. This is primarily attributed to the particular designs for cross domain scenarios of activity recognition. While, among these two methods, our proposed WiTeacher further significantly improves the recognition accuracy. For example, the accuracy of WiTeacher is approximately 2% higher than that of MetaActivity on both datasets. Note that MetaActivity is a meta-learning-based method which requires several labeled samples per category from the target domain, and three samples per category are selected for this experiment. However, our model does not require any labeled data from the target domain, and is highly competitive compared to MetaActivity. The main reason is that, coupled with our designed adaptive labels, target-like samples generated by GANs can efficiently enable the model to capture characteristics of the target data, and further enhance recognition performance.

Second, the Mean Teacher-based method, LCMTeacher, obtain better accuracy than GAN-based approaches such as ManiGAN and CsiGAN. This results indicate that although Mean Teacher is not specially designed for cross domain scenarios, it can take advantage of unlabeled target data to boost the generalization capacity of the model to the target domain. And Mean Teacher is also promising to achieve better performance in semi-supervised learning. This is the reason that we introduce Mean Teacher as the basic framework for cross domain activity recognition. However, the original Mean Teacher model suffers from the problems of being biased towards the source domain and ignoring the relationship between samples. Hence, by incorporating the label smoothing-based classification loss and sample relation-based consistency regularization term, our model WiTeacher remarkably outperforms LCMTeacher.

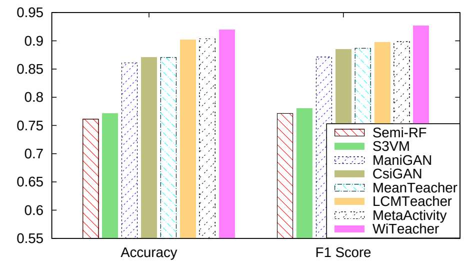


Fig. 2: The activity recognition performance for SignFi data

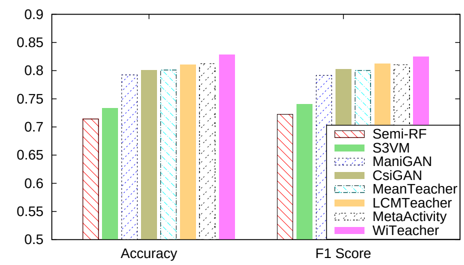


Fig. 3: The activity recognition performance for DeepSeg data

C. Ablation study

There are a few design components within the WiTeacher framework. To investigate the efficacy of these components, we compare its variants from several aspects. We evaluate the model's effectiveness by ablating these components. Specifically, we generate the following variants of WiTeacher:

- **WiT-Base** is a basic Mean Teacher model that remove the classification loss with label smoothing and the consistency loss of sample relation.
- **WiT-GAN** is a model that incorporates the classification loss based on target-like samples generated by StyleGAN into the WiT-Base model but without label smoothing.
- **WiT-LS** is a model that involves the classification loss with label smoothing for the WiT-GAN model.
- **WiT-SR** is a model that takes into account the consistency loss of sample relation on the basis of the WiT-Base model.
- **WiT-Full** is the full model that considers both the classification loss with label smoothing and the consistency loss of sample relation.

TABLE I: The performance of different design choices

	DataSet	WiT-Base	WiT-GAN	WiT-LS	WiT-SR	WiT-Full
SignFi	Accuracy	87.04	89.11	91.06	89.13	**91.94**
	F1 Score	88.69	90.12	91.42	90.26	**92.63**
DeepSeg	Accuracy	80.11	81.09	82.04	81.14	**82.86**
	F1 Score	80.02	80.67	81.46	80.93	**82.43**

The experimental results using SignFi data and DeepSeg data are presented in Table I with the best results highlighted in boldface. As seen, when incorporating the classification loss based on target-like samples generated by StyleGAN but without label smoothing, *WiT-GAN* apparently outperforms *WiT-Base* on both datasets. Further, *WiT-LS*, which takes the proposed label smoothing method into account, achieves a significant improvement compared to *WiT-GAN*. This indicates

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3256324

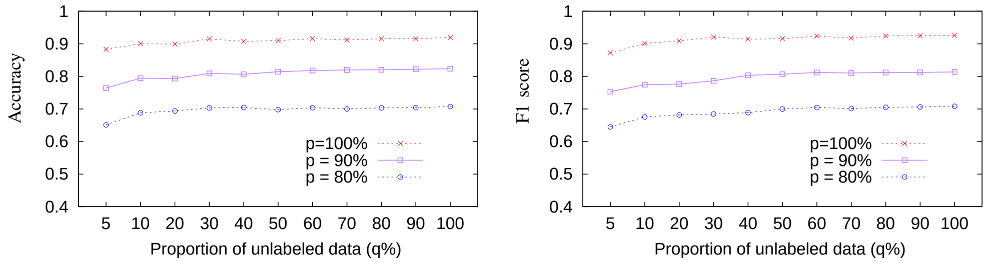


Fig. 4: The performance with different size of unlabeled data for SignFi data

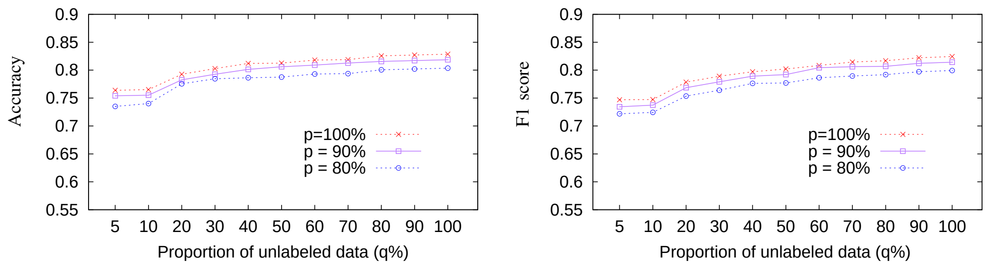


Fig. 5: The performance with different size of unlabeled data for DeepSeg data

that our designed label smoothing method can effectively contribute to the promotion of activity performance. In addition, by introducing the sample relation-based consistency loss, *WiT-SR* acquires better performance than *WiT-Base*. This result proves the advantages of considering sample relation that can be regarded as a data augmentation method to enhance model performance. However, for these methods of deleting one or tow parts, there also exists distinct performance degradation compared with the model with all the components *WiT-Full*. These results suggest that the components we designed can effectively facilitate the recognition performance.

D. Impact of Unlabeled Data Size

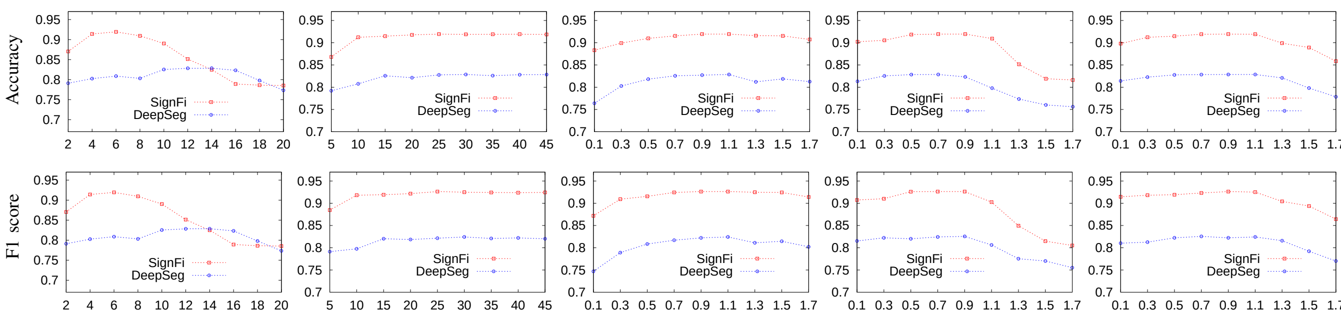
As a kind of semi-supervised method, our model WiTeacher mainly takes advantage of unlabeled target data to enhance the generalization capacity. Here we investigate the impact of the number of unlabeled target data on recognition performance by conducting semi-supervised classification with different proportions of unlabeled data. In this experiment, we select p = [80, 90, 100]% of all the labeled samples from the source domain as labeled data, and select q% of all the unlabeled samples from the target domain as unlabeled data. When q = 100%, the experiments are fully semi-supervised learning.

Figure 4 and Figure 5 present the recognition accuracy and F1 score for SignFi data and DeepSeg data. As shown, the performance of WiTeacher increases along with the grow of the unlabeled data size, suggesting that this size plays an important role on activity recognition results. However, the performance becomes relatively stable when the proportion of unlabeled data reaches 30% for SignFi data, and 40% for DeepSeg data. Specifically, for SignFi data, the accuracy at $q = 30\%$ reaches 91.51%, which is quite close to the accuracy (91.59%) at $q = 90\%$. For DeepSeg data, the accuracies at $q = 40\%$ and $q = 90\%$ is very similar. This indicates that WiTeacher can effectively adopt a part of unlabeled data to obtain a great improvement on recognition performance. In addition, it is apparent that the performance of p = 100% is always higher than that of other p values, which illustrates that the labeled source data can also boost the model performance, and adequate labeled source data is required for better performance because it needs a certain amount of labeled data to train both mean teacher model and StyleGAN. However, the source data can be collected in a specific environment and annotated in a comparatively easy way. In addition, the accuracy gap between p = 100% and p = 90% of SignFi data is obviously greater than that of DeepSeg data. This is because the number of labeled data per category has a great difference for these two datasets.

E. Parameter sensitivity

We here observe the role of the important parameters in the WiTeacher framework, i.e., the number of the top similar samples Z in Equation 5, the number of selected sample pairs F in Equation 9 and the three weight parameters, λ_1, λ_2 and λ_3, in Equation 10. Figure 6 presents the parameter sensitivity analysis in terms of the two metrics across the two datasets.

Figure 6(a) illustrates that the recognition performance of WiTeacher initially rises along with the increase of Z values before its peak and subsequently declines. When the Z value is extremely small, only very few of the most similar samples are selected to compute the distance D^i_{target}. This can result in a smaller distance, and further a higher quality score, which indicates all the generated samples are like a real one, and the soft labels computed by Equation 6 will approach the hard labels of the source data. Hence, the soft labels computed using a small Z value cannot accurately represent category information of the generated samples with noises and distortions, which leads to performance degradation. At the same time, a bigger Z value also result in performance decline because too many samples may exceed the number of samples in the corresponding category, and induce noises for computing the quality score. Also, the peaks of the two datasets are located at different Z values since the activity numbers per category in these two datasets are different.



(a) Z (b) F (c) λ_1 (d) λ_2 (e) λ_3

Fig. 6: The influence of different parameters.

Figure 6(b) presents the role of the number of selected sample pairs F in Equation 9. The x-axis refers to the ratio of the selected pairs among all the pairs. Here N samples can generate $N \cdot (N-1)$ sample pairs. This figure shows that 10% of sample pairs for SignFi data and 15% for DeepSeg data are adequate for WiTeacher to obtain better accuracy. The different proportions for both datasets are attributed to the different number of unlabeled data in these two datasets. This result indicates that the number of selected sample pairs exert an important role in recognition performance. And too few sample pairs cannot display the advantages of the sample relation-based consistency loss. At the same time, more sample pairs are unnecessary for activity recognition.

The impact of the three parameters, λ_1, λ_2 and λ_3, in Equation 10 are presented in Figure 6(c), 6(d) and 6(e), respectively. For λ_1, as shown in Figure 6(c), both smaller value and bigger value result in a lower accuracy. The reason is that a smaller λ_1 can weaken the function of the standard consistency loss $\mathcal{L}_{unlab}$ (which enables the model to learn the characteristics of target data) and further lead to lower performance. On the other hand, a bigger λ_1 might attenuate the effectiveness of other losses, such as $\mathcal{L}_{lab}$, which may deteriorate overall recognition performance. For the same reason, the impact of λ_2 and λ_3, which are presented in Figure6(d) and 6(e), exhibits the similar trends.

V. Related Work

Here we describe the relevant researches from two aspects, including CSI-based activity recognition and Mean Teacher-based semi-supervised learning. And the main differences of our study and these researches are highlighted.

A. CSI-based activity recognition

According to the application scenario, CSI-based activity recognition works fall into two folds: recognizing activities under the same scenarios and different scenarios. Some researches focus on improving recognition performance under the scenarios that both training data and test data are from the same subjects and environments. For instance, Khan *et al.* [46] proposed to adopt differential CSI to alleviate the offset and background noise, and then use a LSTM model to recognize human behaviour in a fixed IoT environment. Guo *et al.* [47] built an air handwriting recognition system using WiFi CSI, WiReader, which combines principal component analysis and discrete wavelet transform to extract a correlation feature matrix, and then complete handwriting classification using L-STM. Despite of being able to apply to different scenarios, this model is designed and performs well only for the situation that the training and test data are from the same scenario. Cui *et al.* [48] explored the spatial-correlation between different CSI subcarriers, and adopted diversified deep ensemble methods for single-user activity recognition. Zou *et al.* [49] presented a multiple kernel semi-representation learning method, MKSRL, which allows the input of expert domain knowledge in a flexible way and conducts automatic and effective multi-kernel representation learning for the activity recognition task. Sheng *et al.* [50] designed a deep spatial-temporal learning framework, which concentrating on how to extract features with spatial and time-dependent information for CSI-based action recognition. Wu *et al.* [51] revised an opposite robust PCA to process the CSI data and then designed a normalized variance sliding window approach to segment activities for the situation of the WiFi signal across the wall. Wang *et al.* [52] built a human activity recognition system, CARM, which estimates the correlation between CSI dynamics and human activities, and recognizes a given activity by comparing this correlation.

Other works aim at conducting cross-domain activity recognition to address the performance decline problem when test scenarios (target domains) are different from training ones (source domains). One way of addressing this performance degradation is to design few-shot learning-based solutions, and the main idea of which is to pre-train a model based on data from the source domain, and then identify activities based on a few labeled samples from the target domain. For example, Yang *et al.* [53] presented a deep Siamese representation learning architecture for one-shot gesture recognition, which adopts the Siamese framework and transferable pairwise loss to alleviate the problem of environmental dynamics and individual heterogeneity. Sheng *et al.* [54] integrated Bidirectional LSTM into CNN to obtain the spatial-temporal features of CSI for cross-scene action recognition. They take an off-the-shelf model as the pre-trained model and then fine-tune it in the new scenario to solve the problem that the trained model fully fails with environmental changes. Zhang *et al.* [9] designed an adaptable CSI activity recognition system based on meta-learning, which updates the pre-trained model through one or more gradient steps with a small amount of labeled samples from new environments to implement cross scene activity recognition.

Another way of dealing with cross-domain activity recognition is to propose GAN-based data augmentation methods, which introduce a generator to produce diverse samples for improving classifier robustness. For instance, Zou *et al.* [14] built a WiFi-enabled device-free adaptive gesture recognition scheme, WiADG, which introduces a generator to map the unlabeled target data to a domain invariant latent feature space for reducing the domain discrepancy between the source and the target domain. Jiang *et al.* [55] designed a deep-learning based device-free activity recognition framework, which can extract environment and individual-independent features shared by the collected activity data on different individuals under different environments. Xiao *et al.* [13] proposed a semi-supervised GAN for CSI-based activity recognition, CsiGAN, which introduces a new complement generator for GANs to produce diverse fake samples for training a robust classifier. Wang *et al.* [12] proposed a multimodal generator to approximate the CSI data distribution in different environment settings with limited measured CSI data, which can improve model robustness under multiple unexpected dynamic changes within the environment.

In spite of obtaining substantial improvement for cross domain activity recognition, the above works still have some shortages. For example, few-shot learning-based methods also require a few labeled samples from the target domain. However, it is still hard to collect these samples, especially for older adults. GAN-based data augmentation methods generally treat all generated samples equally. However, due to unstable nature of GANs, some generated samples might approach the real one, while others might be quite noised. Our model is a kind of GAN-based data augmentation methods, which does not require any labeled data from the target domain. Meanwhile, we design a label smoothing-based method to produce proper soft labels, which can more accurately represent category information of generated samples. Also we design a sample relation-based consistency regularization term to enhance model performance.

B. Mean Teacher-based semi-supervised learning

By using consistency regularization in semi-supervised deep learning, Mean Teacher-based methods achieve state-of-the-art results in the fields of image processing. These works try to address the problem of limited labeled data from the target domain, and mainly focus on image classification, object detection, and image segmentation. In the area of image classification, French *et al.* [56] developed an effective domain adaptation algorithm, which can achieve excellent classification results in image domain adaptation benchmarks. Guo *et al.* [57] developed a novel uncertainty filter based on Mean Teacher, which selects reliable unlabeled data for initial training steps to enhance detection performance in facade inspection. Li *et al.* [58] built a master-teacher-student approach, where the master network integrates the knowledge of the student and teacher models with additional access to a few newly discovered samples. Ye *et al.* [59] presented a novel approach to reduce the classifier bias to source samples for unsupervised domain adaptation by matching the distribution of the two domains.

For object detection, Cai *et al.* [60] presented a Mean Teacher-based method that remolds the model under the framework of Faster R-CNN by combining the object relations into the measure of consistency cost between teacher and student models. Chen *et al.* [61] assigned a multitasking module to the student and teacher networks in the mean teacher framework to exploit auxiliary unlabeled data to enhance the shadow recognition accuracy. Xiong *et al.* [62] proposed SOAP, a new source data-free domain adaptation method through domain perturbation, where the Mean Teacher model is used to optimize three consistency regularization terms by aligning the source domain to the target one. Deng *et al.* [63] designed a cross-domain distillation method based on Mean Teacher to explore the expert knowledge from the teacher model in cross-domain object detection. For image segmentation, Li *et al.* [64] presented a novel regularized Mean Teacher method whose student model is optimized by multi-scale deep supervision and hierarchical consistency regularization for 3D Left Atrium Segmentation. Xie *et al.* [34] designed a confidence framework to forecast the model confidence guided by the true class probability, and then the student model extracts trustworthy targets from the teacher model to enhance the results of skin lesions segmentations.

The above researches mainly focus on the fields of image processing. Instead, we apply Mean Teacher to WiFi CSI-based human activity recognition, and design an adaptive label smoothing method and a sample relation-based consistency regularization for time-series data, which are incorporated into the Mean Teacher model for improving recognition performance.

VI. Conclusions

In this paper, we presented WiTeacher, a Mean Teacher-based cross-domain human activity recognition framework using WiFi CSI. In this framework, we designed an adaptive label smoothing method to produce proper soft labels for target-like samples generated by StyleGAN. Based on these target-like samples with soft labels, we built a label smoothing-based classification loss to promote the generalization capacity of the model. Further, we presented a sample relation-based consistency regularization term to force the distance of two samples to be consistent with the augmented ones, which can make the model more robust. Through experimental evaluations on the two public datasets, we illustrate that WiTeacher dramatically improves activity recognition performance and outperforms state-of-the-art baselines. In the future, we plan to study the feasibility in the context of different deployed scenarios, especially for cross-scene applications where the training data and test data are collected in different environments. Also, we will apply this model to activity segmentation and devise a real-time activity recognition system for healthcare services.

REFERENCES

[1] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–40, 2021.

[2] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, p. 107561, 2020.

[3] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Computing Surveys*, vol. 52, no. 3, pp. 1–36, 2019.

[4] Z. Hussain, Q. Z. Sheng, and W. E. Zhang, "A review and categorization of techniques on device-free human activity recognition," *Journal of Network and Computer Applications*, vol. 167, p. 102738, 2020.

[5] R. Fallahzadeh, Z. E. Ashari, P. Alinia, and H. Ghasemzadeh, "Personalized activity recognition using partially available target data," *IEEE Transactions on Mobile Computing*, vol. 22, no. 1, pp. 374–388, 2023.

[6] Z. Shi, J. A. Zhang, R. Y. Xu, and Q. Cheng, "Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning," *IEEE Transactions on Mobile Computing*, vol. 21, no. 2, pp. 540–554, 2022.

[7] T. Gong, Y. Kim, R. Choi, J. Shin, and S.-J. Lee, "Adapting to unknown conditions in learning-based mobile sensing," *IEEE Transactions on Mobile Computing*, vol. 21, no. 10, pp. 3470–3485, 2021.

[8] Z. Wang, B. Guo, Z. Yu, and X. Zhou, "Wi-fi csi-based behavior recognition: From signals and actions to activities," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 109–115, 2018.

[9] Y. Zhang, X. Wang, Y. Wang, and H. Chen, "Human activity recognition across scenes and categories based on csi," *IEEE Transactions on Mobile Computing*, vol. 21, no. 7, pp. 2411–2420, 2022.

[10] S. Ding, Z. Chen, T. Zheng, and J. Luo, "Rf-net: a unified meta-learning framework for rf-enabled one-shot human activity recognition," in *Proceedings of the Conference on Embedded Networked Sensor Systems*, 2020, pp. 517–530.

[11] Y. Zhang, Y. Chen, Y. Wang, Q. Liu, and A. Cheng, "Csi-based human activity recognition with graph few-shot learning," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4139–4151, 2021.

[12] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Multimodal csi-based human activity recognition using gans," *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17 345–17 355, 2021.

[13] C. Xiao, D. Han, Y. Ma, and Z. Qin, "Csigan: Robust channel state information-based activity recognition with gans," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 191–10 204, 2019.

[14] H. Zou, J. Yang, Y. Zhou, L. Xie, and C. J. Spanos, "Robust wifi-enabled device-free gesture recognition via unsupervised adversarial domain adaptation," in *The International Conference on Computer Communication and Networks*, 2018, pp. 1–8.

[15] Y. Wang, L. Yao, Y. Wang, and Y. Zhang, "Robust csi-based human activity recognition with augment few shot learning," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24 297–24 308, 2021.

[16] C. Xiao, S. Chen, F. Zhou, and J. Wu, "Self-supervised few-shot time-series segmentation for activity recognition," *IEEE Transactions on Mobile Computing*, pp. 1–14, 2022.

[17] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8110–8119.

[18] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," in *International conference on machine learning*, 2018, pp. 1989–1998.

[19] Q. Yang, X. Wei, B. Wang, X.-S. Hua, and L. Zhang, "Interactive self-training with mean teachers for semi-supervised object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 5941–5950.

[20] J. Deng, W. Li, Y. Chen, and L. Duan, "Unbiased mean teacher for cross-domain object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4091–4101.

[21] Y. Li, H. Yao, and C. Xu, "Test: Triplet ensemble student-teacher model for unsupervised person re-identification," *IEEE Transactions on Image Processing*, vol. 30, pp. 7952–7963, 2021.

[22] X. Liu and S. Zhang, "Graph consistency based mean-teaching for unsupervised domain adaptive person re-identification," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2021, pp. 874–880.

[23] 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*, June 2019, pp. 4401–4410.

[24] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 12 104–12 114.

[25] Y. Luo, J. Zhu, M. Li, Y. Ren, and B. Zhang, "Smooth neighbors on teacher graphs for semi-supervised learning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8896–8905.

[26] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 1195–1204.

[27] R. Muller, S. Kornblith, and G. Hinton, "When does label smoothing help?" in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019, p. 422.

[28] W. Li, G. Dasarathy, and V. Berisha, "Regularization via structural label smoothing," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, vol. 108, 2020, pp. 1453–1463.

[29] Y. Ma, Y. Lei, and T. Wang, "A natural scene recognition learning based on label correlation," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 1, pp. 150–158, 2022.

[30] B. Guo, S. Han, X. Han, H. Huang, and T. Lu, "Label confusion learning to enhance text classification models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 12 929–12 936.

[31] M. Lukasik, H. Jain, A. Menon, S. Kim, S. Bhojanapalli, F. Yu, and S. Kumar, "Semantic label smoothing for sequence to sequence problems," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Nov. 2020, pp. 4992–4998.

[32] Q. Wang, W. Min, Q. Han, Q. Liu, C. Zha, H. Zhao, and Z. Wei, "Inter-domain adaptation label for data augmentation in vehicle re-identification," *IEEE Transactions on Multimedia*, vol. 24, pp. 1031–1041, 2022.

[33] Z. Zhu, X. Jiang, F. Zheng, X. Guo, F. Huang, X. Sun, and W. Zheng, "Viewpoint-aware loss with angular regularization for person re-identification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 13 114–13 121, 2020.

[34] Z. Xie, E. Tu, H. Zheng, Y. Gu, and J. Yang, "Semi-supervised skin lesion segmentation with learning model confidence," in *International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 1135–1139.

[35] Z. H. Chen, L. Zhu, L. Wan, S. Wang, W. Feng, and P. A. Heng, "A multi-task mean teacher for semi-supervised shadow detection," in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 2020, pp. 5611–5620.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, 2015.

[37] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi: Sign language recognition using wifi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–21, 2018.

[38] C. Xiao, Y. Lei, Y. Ma, F. Zhou, and Z. Qin, "Deepseg: Deep-learning-based activity segmentation framework for activity recognition using wifi," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5669–5681, 2021.

[39] S. Ding, Z. Zhu, and X. Zhang, "An overview on semi-supervised support vector machine," *Neural Computing and Applications*, vol. 28, no. 5, pp. 969–978, 2017.

[40] I. Triguero, S. García, and F. Herrera, "Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study," *Knowledge and Information Systems*, vol. 42, no. 2, pp. 245–284, 2015.

[41] B. Lecouat, C. S. Foo, H. Zenati, and V. R. Chandrasekhar, "Semi-supervised learning with gans: Revisiting manifold regularization," in *Proceedings of the International Conference on Learning Representations*, 2018.

[42] Z. Chen, B. Dutton, B. Ramachandra, T. Wu, and R. R. Vatsavai, "Local clustering with mean teacher for semi-supervised learning," *The 25th International Conference on Pattern Recognition*, pp. 6243–6250, 2020.

[43] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "Falldefi: Ubiquitous fall detection using commodity wi-fi devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–25, 2018.

[44] G.-J. Qi, L. Zhang, H. Hu, M. Edraki, J. Wang, and X.-S. Hua, "Global versus localized generative adversarial nets," in *The IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1517–1525.

[45] T. Miyato, S.-I. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: A regularization method for supervised and semi-supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1979–1993, 2019.

[46] P. Khan, B. S. K. Reddy, A. Pandey, S. Kumar, and M. Youssef, "Differential channel-state-information-based human activity recognition in iot networks," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11 290–11 302, 2020.

[47] Z. Guo, F. Xiao, B. Sheng, H. Fei, and S. Yu, "Wireader: Adaptive air handwriting recognition based on commercial wifi signal," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10 483–10 494, 2020.

[48] C. A. Wei, L. B. Bing, Z. A. Le, and A. Zc, "Device-free single-user activity recognition using diversified deep ensemble learning," *Applied Soft Computing*, vol. 102, p. 107066, 2021.

[49] H. Zou, Y. Zhou, R. Arghandeh, and C. J. Spanos, "Multiple kernel semi-representation learning with its application to device-free human activity recognition," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7670–7680, 2019.

[50] B. Sheng, Y. Fang, F. Xiao, and L. Sun, "An accurate device-free action recognition system using two-stream network," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7930–7939, 2020.

[51] X. Wu, Z. Chu, P. Yang, C. Xiang, X. Zheng, and W. Huang, "Tw-see: Human activity recognition through the wall with commodity wi-fi devices," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 306–319, 2019.

[52] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial wifi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[53] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from wifi: A siamese recurrent convolutional architecture," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 763–10 772, 2019.

[54] B. Sheng, F. Xiao, L. Sha, and L. Sun, "Deep spatialctemporal model based cross-scene action recognition using commodity wifi," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3592–3601, 2020.

[55] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *Proceedings of the Annual International Conference on Mobile Computing and Networking*, 2018, pp. 289–304.

[56] G. French, M. Mackiewicz, and M. Fisher, "Self-ensembling for visual domain adaptation," in *International Conference on Learning Representations*, 2018.

[57] J. Guo, Q. Wang, and Y. Li, "Semi-supervised learning based on convolutional neural network and uncertainty filter for facade defects classification," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 3, pp. 302–317, 2021.

[58] Y. Li, Z. Zhao, H. Sun, Y. Cen, and Z. He, "Snowball: Iterative model evolution and confident sample discovery for semi-supervised learning on very small labeled datasets," *IEEE Transactions on Multimedia*, vol. 23, pp. 1354–1366, 2021.

[59] Y. Ye, Z. Huang, T. Pan, J. Li, and H. T. Shen, "Reducing bias to source samples for unsupervised domain adaptation," *Neural Networks*, vol. 141, pp. 61–71, 2021.

[60] Q. Cai, Y. Pan, C.-W. Ngo, X. Tian, L. Duan, and T. Yao, "Exploring object relation in mean teacher for cross-domain detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 457–11 466.

[61] Z. Chen, L. Zhu, L. Wan, S. Wang, W. Feng, and P.-A. Heng, "A multi-task mean teacher for semi-supervised shadow detection," in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 2020, pp. 5611–5620.

[62] L. Xiong, M. Ye, D. Zhang, Y. Gan, X. Li, and Y. Zhu, "Source data-free domain adaptation of object detector through domain-specific perturbation," *International Journal of Intelligent Systems*, vol. 36, no. 8, pp. 3746–3766, 2021.

[63] J. Deng, W. Li, Y. Chen, and L. Duan, "Unbiased mean teacher for cross-domain object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4091–4101.

[64] S. Li, Z. Zhao, K. Xu, Z. Zeng, and C. Guan, "Hierarchical consistency regularized mean teacher for semi-supervised 3d left atrium segmentation," in *Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, 2021, pp. 3395–3398.

Chunjing Xiao received the Ph.D. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2013.

He is currently an Associate Professor with the School of Computer and Information Engineering, Henan University, Kaifeng, China. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. His current research interests include Internet of Things, anomaly detection, and recommender systems.

Yue Lei received the B.E degree from the School of Computer and Information Engineering, Henan University, China, in 2019. He is currently pursuing the M.S. degree in the School of Computer and Information Engineering, Henan University.

His current research interests include Internet of Things, anomaly detection and data analytics.

Chun Liu received the Ph.D. degree in software engineering from the Institute of Mathematics, Chinese Academy of Sciences, Beijing, China, in 2012.

From 2012 to 2016, he was a Lecturer with the School of Computer and Information Engineering, Henan University, Kaifeng, China. From 2014 to 2015, he was a Visiting Scholar with the School of Computing and Communications, The Open University, Milton Keynes, U.K. Since 2016, he has been an Assistant Professor with the Henan Key Laboratory of Big Data Analysis and Processing and the Henan Industrial Technology Academy of Spatio-Temporal Big Data, Henan University. His research interests include software engineering and big data.

Jie Wu received the Ph.D. degree in computer engineering from Florida Atlantic University, Boca Raton, FL, USA, in 1989.

He is the Director of the Center for Networked Computing and Laura H. Carnell professor at Temple University. He also serves as the Director of International Affairs at College of Science and Technology. He served as Chair of Department of Computer and Information Sciences from the summer of 2009 to the summer of 2016 and Associate Vice Provost for International Affairs from the fall of 2015 to the summer of 2017. Prior to joining Temple University, he was a program director at the National Science Foundation and was a distinguished professor at Florida Atlantic University. His current research interests include mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications. Dr. Wu regularly publishes in scholarly journals, conference proceedings, and books. He serves on several editorial boards, including IEEE Transactions on Mobile Computing, IEEE Transactions on Service Computing, Journal of Parallel and Distributed Computing, and Journal of Computer Science and Technology. Dr. Wu was general co-chair for IEEE MASS 2006, IEEE IPDPS 2008, IEEE ICDCS 2013, ACM MobiHoc 2014, ICPP 2016, and IEEE CNS 2016, as well as program co-chair for IEEE INFOCOM 2011 and CCF CNCC 2013. He was an IEEE Computer Society Distinguished Visitor, ACM Distinguished Speaker, and chair for the IEEE Technical Committee on Distributed Processing (TCDP). Dr. Wu is a Fellow of the AAAS and a Fellow of the IEEE. He is the recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award.