Democratizing Federated WiFi-based Human Activity Recognition Using Hypothesis Transfer

Bing Li, Wei Cui, Le Zhang, Qi Yang, Min Wu, Joey Tianyi Zhou, *Senior Member, IEEE*

***Abstract*—Human activity recognition (HAR) is a crucial task in IoT systems with applications ranging from surveillance and intruder detection to home automation and more. Recently, non-invasive HAR utilizing WiFi signals has gained considerable attention due to advancements in ubiquitous WiFi technologies. However, recent studies have revealed significant privacy risks associated with WiFi signals, raising concerns about bio-information leakage. To address these concerns, the decentralized paradigm, particularly federated learning (FL), has emerged as a promising approach for training HAR models while preserving data privacy. Nevertheless, FL models may struggle in end-user environments due to substantial domain discrepancies between the source training data and the target end-user environment. This discrepancy arises from the sensitivity of WiFi signals to environmental changes, resulting in notable domain shifts. As a consequence, FL-based HAR approaches often face challenges when deployed in real-world WiFi environments. Albeit there are pioneer attempts on federated domain adaptation, they typically require non-trivial communication and computation cost, which is prohibitively expensive especially considering edge-based hardware equipment of end-user environment. In this paper, we propose a model to democratize the WiFi-based HAR system by enhancing recognition accuracy in unannotated end-user environments while prioritizing data privacy. Our model leverages the hypothesis transfer and a lightweight hypothesis ensemble to mitigate negative transfer. We prove a tighter theoretical upper bound compared to existing multi-source federated domain adaptation models. Extensive experiments shows our model improves the average accuracy by approximately 10 absolute percentage points in both cross-person and cross-environment settings comparing several state-of-the-art baselines.**

***Index Terms*—Domain adaptation, Intelligent sensing, Human activity recognition, Federated learning**

I. Introduction

HUman activity recognition (HAR) plays a crucial role in IoT systems, offering extensive applications such as intruder detection, patient monitoring in healthcare facilities, and supervision of children and the elderly in smart homes [1]. In the past decade, researchers have proposed and examined various HAR systems employing different techniques [2]. Wearable sensor-based and vision-based HAR techniques are the most common prototypes [3], [4]. However, both approaches have significant limitations that restrict their real-world applications. Wearable sensor-based systems require individuals to wear cumbersome devices like accelerometers and gyroscopes, which are especially inconvenient for children and the elderly. On the other hand, vision-based HAR systems rely on strict working conditions such as camera deployment, adequate illumination, and unobstructed views (i.e., line-of-sight). In addition, the vision-based systems also raise privacy and security concerns by directly collecting users' bio-information.

In recent years, there has been a surge of interest in non-invasive HAR using WiFi signals, capitalizing on the rapid advancements in ubiquitous WiFi technologies. They operate on the facts that human activities generate unique variations (a.k.a. multi-path and the fading effect) in wireless signals between a transmitter (TX) and a receiver (RX), as depicted in Fig. 1. WiFi-based HAR systems offer distinct advantages. Firstly, they can be easily deployed in an off-the-shelf manner since WiFi infrastructures are widely available in indoor environments. Secondly, they are user-friendly and eliminate the need for individuals to wear any devices. Thirdly, they demonstrate robust performance in various environmental conditions, including non-line-of-sight (NLOS) scenarios and non-illumination conditions.

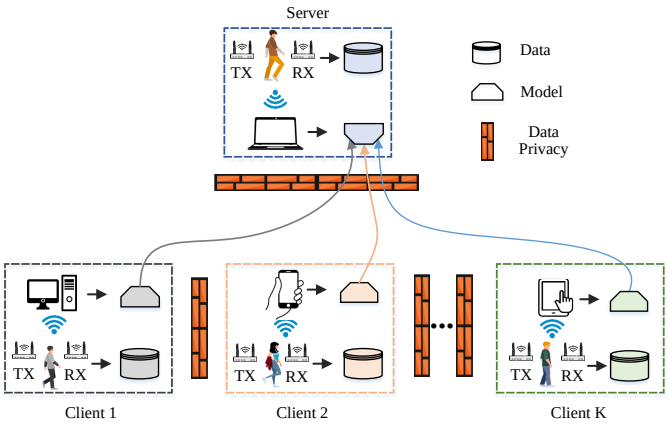


Fig. 1. A typical federated WiFi-based HAR system.

Unfortunately, recent research [5], [6] has revealed a significant vulnerability in WiFi signals, resulting in a substantial risk of user identity leakage. This highlights serious privacy concerns that require immediate attention. In response to this concern, federated WiFi-based HAR approaches [7], [8] have been developed. These approaches typically utilize the federated learning (FL) paradigm [9] to keep data access strictly decentralized to each client. Fig. 1 illustrates a typical

B. Li and L. Zhang are with the University of Electronic Science and Technology of China.

W. Cui and M. Wu are affiliated with Institute for Infocomm Research (I^2R), Agency for Science, Technology and Research (A*STAR), Singapore.

J. T. Zhou is with a. Centre for Frontier AI Research (CFAR), Agency for Science, Technology and Research (A*STAR), Singapore; b. Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research(A*STAR), Singapore.

Q. Yang is with Shandong University of Science and Technology, China.

W. Cui and J. T. Zhou are corresponding authors.

federated WiFi-based HAR system. Each client retains its specific user or environment data, and there is no direct data sharing between clients. Despite FL-based HAR approaches demonstrating satisfactory data privacy and security properties, they still encounter two significant challenges that hinders its real-life applications:

(*i*) The non-i.i.d. data distribution of clients. A prerequisite of the performance convergence of federated learning, as proved by *FedAvg Theorem* [10], is the i.i.d. assumption. However, WiFi signals are highly vulnerable to environmental factors such as space, participants, and transmitter (TX) and receiver (RX) placement – even small changes, such as slightly adjusting antenna angles or positions, can lead to significant shifts in data distribution and a subsequent decline in performance. Facing such highly non-i.i.d. characteristics observed in real-world WiFi environments, the conventional federated learning paradigm may not be adequately suited for such scenarios, due to the canceling effect [11] and negative transfer [12].

(*ii*) Domain shifts when deploying to the end-user environment. Even if the FL model could successfully converge to the centralized model, the vulnerability of WiFi signals to environmental changes clearly indicates a significant domain shift between the source environments and the end user's target environment[1] To ensure satisfactory performance for end users, re-annotation and fine-tuning of the model on the user side are necessary. However, the process of annotating, storing, and training for end users is impractical due to the prohibitively large overhead[2]. Compared to other sensing sources like cameras and Millimeter Wave Radar (MWR), which are minimally affected by domain shifts, WiFi CSI is extremely vulnerable to domain discrepancies due to the sensitivity of WiFi signals to environmental changes. This often results in very low end-user performance when deploying WiFi-based HAR in target end-user environments. Some research resorts to unsupervised multi-source domain adaptation (UMDA) methods [13], [14]. However, UMDA requires direct access to all source domains, which contradicts the decentralized nature of the FL paradigm, where data and computations must be kept distributed. Although recent works have explored decentralized UMDA [15], [16], they often entail non-trivial communication or computational costs and suffer from negative transfer in multi-source scenarios. This poses a particular challenge when considering that many WiFi-based HAR systems are deployed on edge devices [17], such as mobile phones. These factors render them impractical, especially when dealing with limited computational resources in WiFi-based HAR tasks.

In this study, we propose a model working towards democratizing the WiFi-based HAR system by improving recognition accuracy in unannotated end-user target environments, and prioritizing data privacy, while also working efficiently in communication and computation. Our *Federate-lightweight HAR via Hypothesis Transfer* (FligHT) model leverages the hypothesis invariant assumption of WiFi data and consists of three key components: asynchronous training through hypothesis transfer, a lightweight hypothesis ensemble to mitigate negative transfer without large computational costs, and model perturbation to protect client models against deep leakage attacks using differential privacy. Unlike FL methods that focus on handling non-i.i.d. problems, such as ClusterFL [18], a major challenge of our task is to achieve unsupervised domain adaptation within a federated paradigm. Our novel hypothesis transfer method is specifically tailored to the WiFi-based HAR task, considering the unique characteristics of WiFi signals. We prove a tighter theoretical upper bound for FligHT compared to existing multi-source federated domain adaptation models. Extensive experiments conducted on our realistic end-user emulation platform demonstrate remarkable performance improvements in cross-person and cross-environment scenarios, surpassing the state-of-the-art model by a significant margin of over 10 percentage points.

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

- To the best of our knowledge, we make the first effort to consider a federated unsupervised domain adaptation setting in the WiFi-based HAR task. In this setting, our primary goal is to enhance recognition accuracy in unannotated end-user environments by minimizing domain discrepancies while ensuring data privacy between clients and servers.
- We propose a model for this WiFi-based HAR setting, built upon the assumption of *Hypothesis Invariance of WiFi Pattern*, that addresses the high non-i.i.d. patterns and domain discrepancies of WiFi signals.
- We prove a tighter theoretical upper bound for our model compared to state-of-the-art multi-source federated domain adaptation models, ensuring better generalization performance in multi-source settings.
- We conduct extensive experiments on both self-collected data using our platform and an open-source dataset. The results demonstrate remarkable performance improvements in cross-person and cross-environment scenarios.

II. Related Works

A. WiFi-based Human Activity Recognition

WiFi-based human activity recognition (HAR) has been extensively studied, utilizing techniques such as signal preprocessing, feature extraction, and machine learning classifiers [2]. Recently, deep learning methods, including bidirectional LSTM (BLSTM) [19], Trasformer [20], TCN [21], and Autoencoder [22], [23] have been proposed for activity recognition using WiFi channel state information (CSI) signals. However, most existing WiFi-based HAR systems are designed for fixed and controlled environments, leading to performance degradation when deployed in new environments with different spatial layouts and objects. In order to adapt to different environments, several models have been proposed for cross-environment HAR [7], [24] or localization [25]. They typically utilize discrepancy-based domain adaptation approaches (e.g., MMD [26]), or adversarial-based domain adaptation models

[1]Without domain adaptation, a centralized model only achieves a target accuracy of 39.32% in the cross-environment setting, which is far lower than the 74.48% achieved by our model.

[2]Achieving an accuracy of 90% needs 4,000 training instances (2.3 GB).

(e.g., InverGAN [27]). In light of recent research findings [5], [6], it is evident that WiFi signals pose significant privacy risks, potentially leading to user identity leakage. To safeguard user bio-information, it becomes imperative to establish privacy and security measures, aligning with regulations such as the EU's General Data Protection Regulation (GDPR) [28].

B. Federated Learning

Federated learning [29], [30] is a distributed machine learning strategy, it trains a global model by aggregating the local models to approach the centralized datasets. FedAvg [10] has emerged as the standard approach for federated learning. It involves four steps in each round: the server sends a global model to the parties, the parties perform local stochastic gradient descent (SGD) to update their models, the local models are sent back to the server, and finally, the server averages the model weights to produce a global model for the next round. Improvements to FedAvg on non-i.i.d. data can be categorized into local training and aggregation strategies. FedProx [31] introduces a proximal term into the objective function during local training, restricting the local model update by considering the L2-norm distance between the current global model and the local model. Scaffold [32] corrects local updates by incorporating control variates that are updated by each party during local training. The discrepancy between the local and global control variates is used to correct the gradients in local training. FedProto [33] employs a novel prototype learning method to handle the heterogeneous data distributions across clients. Another approach to addressing the non-i.i.d. problem involves federated clustering models. ClusterFL [18] introduces a multi-task federated learning framework that minimizes the empirical training loss across multiple models, effectively capturing the intrinsic clustering relationships among clients. FedCLAR [34] applies federated clustering to sensor-based HAR, grouping users based on the similarity of their local model updates. FedCHAR-DC [35] employs dynamic clustering that adapts to new users or changes in datasets, making it suitable for realistic federated HAR scenarios. Recently, there are some federated WiFi-based HAR approaches [7], [8], [36], which utilize federated learning techniques to keep data access strictly decentralized to each client. However, federated learning methods primarily aim for training convergence in decentralized settings, they may have limitations when deploying WiFi-based HAR systems in diverse end-user environments due to domain shifts.

C. Decentralized Domain Adaptation

To address the domain shift between source and target, conventional domain adaption solutions, including discrepancy-based (e.g., MMD [26]), adversarial-based (e.g., InverGAN [27]), and reconstruction-based methods (e.g., ED [37]) require a direct alignment on features $g(X)$, which has a direct risk of data leakage as being amenable to infer data distribution from features. This is contrary to the user's privacy. FedHAR [7] and similar approaches employ techniques like the Johnson-Lindenstrauss transform, using Laplacian noise, to protect against data re-identification through differential privacy. However, these methods often suffer from a significant performance drop due to feature distortion.

Recently, there are also few works focusing on unsupervised domain adaptation under decentralized settings. FADA [15] introduces the concept of federated domain adaptation and utilizes adversarial training to optimize the H-divergence without accessing data. However, FADA is associated with high communication and computational costs due to sharing output features from the sources for the adversarial training, and it is vulnerable to privacy leakage attacks. Model Adaptation [38] and SHOT [39], [40] are source-free models for decentralized domain adaptation in single-source scenarios. However, they differ from our approach in two ways: *i*) they support only one-shot [41] and do not accommodate multiple communication rounds, and *ii*) they are designed for single-source scenarios and may encounter negative transfer challenges when applied to multi-source situations. Another approach, KD3A [16] makes decentralized domain adaptation by re-training all client models through knowledge distillation on a curated synthetic dataset. However, constructing such curated datasets requires synchronous involvement of all clients, and re-training all client models incurs substantial computational overheads, making it impractical in real WiFi HAR scenarios where parts of clients may be offline and computational resources are limited.

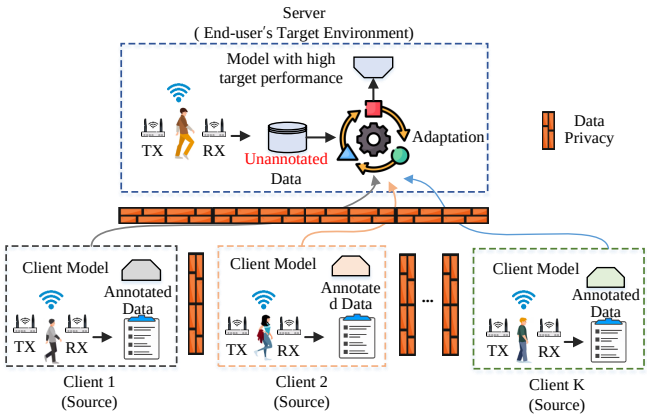


Fig. 2. Overview of the entire process in our problem setting.

III. Methodology

In this section, we present our methodology for democratizing the WiFi-based HAR system. Firstly, we introduce the problem definition and overview of the system. Next, we present the details of client model training and key technical components, including hypothesis transfer, hypothesis ensemble, and model perturbation, along with the overall algorithm. Finally, we provide theoretical insights with a tighter risk upper bound of our method and analyze its communication and computation costs.

A. Problem Setting

1) Problem Definition: Let $\mathbb{D}_s$ and $\mathbb{D}_t$ represent datasets containing WiFi Channel State Information (CSI) from the source and target environments, respectively. We consider

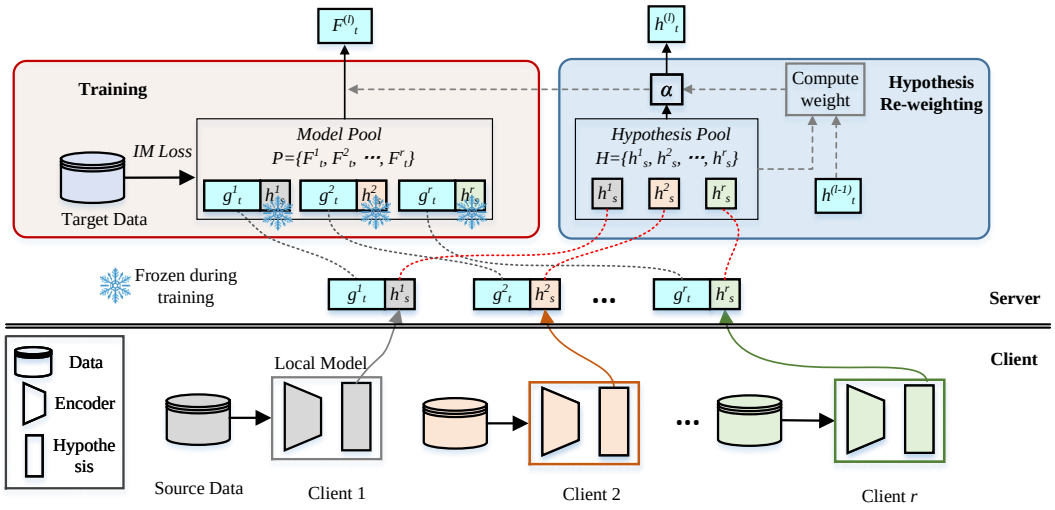


Fig. 3. The framework of our FligHT model.

$K(K \geq 1)$ source environments denoted as $\{\mathbb{D}_s^k\}_{k=1}^K$. Each source dataset contains *annotated* samples $\mathbb{D}_s^k = \{(X_i^k, y_i^k)\}_{i=1}^{N_k}$. The target environment data is denoted as $\mathbb{D}_t = \{X_i^t\}_{i=1}^{N_t}$, which is *unannotated*. Particularly, we consider an identical label space for both the source and target environments consisting of C categories. In particular, aligning with federated settings, the sources act as clients and the target environment acts as the server.

Our primary objective is to learn a model F that minimizes the generalization error $\epsilon_{\mathcal{T}}$ in target environment $\mathcal{T}$, defined as $\epsilon_{\mathcal{T}}(h) = Pr_{(X,y)\sim\mathcal{T}}[h(X) \neq y]$, adhering to the constraint of decentralized data and features access, i.e., *data and features* are strictly kept decentralized to preserve users' bio-information privacy. This constraint ensures that not only data but also features are kept from being shared between the source and target environments. Notably, the prohibition on feature sharing sets us apart from existing approaches in decentralized UMDA, like FADA [15], which are susceptible to model inverse attacks, as identified in studies by Fredrikson et al. [42] and Zhang et al. [43].

In addition to achieving the above objective, we also prioritize the efficiency of the model, ensuring that it is trained without incurring excessive computation and communication overheads.

It is worth noting that our problem definition is fundamentally different from a typical federated learning setting, as ours prioritizes unsupervised target performance, i.e., enhancing recognition accuracy in unannotated end-user environments. Under this setting, handling domain discrepancies while ensuring data privacy is the major challenge. In contrast, the aim of federated learning is merely to learn a model in a distributed manner that is similar to a centralized model, without addressing the unsupervised domain discrepancies problem.

2) Process Overview: The entire process in our problem setting is depicted in Fig. 2. At the client-side (i.e., source environment), the model is trained locally using annotated data. At the server-side (i.e., end-user's target environment), the model is generated by a dynamic adaptation process using the unannotated data of the end-user and client models sent to the server. This process ensures high target performance while maintaining the decentralization of both data and features to protect users' bio-information privacy.

An important advantage of our model is that it doesn't necessitate additional WiFi-related signal processing steps. Many existing techniques involve intricate WiFi signal processing that isn't trainable and operates within a cumbersome pipeline paradigm. Our end-to-end approach provides greater flexibility and simplicity, as validated by our experimental results showcasing the superiority of our model over existing methods.

B. System Overview

In contrast to previous studies, our work aims to bring data privacy and better target performance together. We prioritize data privacy of client users *without aggregating the datasets*, meanwhile, effectively utilized client data to elicit better performance on the non-i.i.d. target environment *without data annotation*. A systematic overview of the proposed FligHT method is provided in Fig. 3. We assume a model can be decomposed into two sequentially stacking parts, i.e., an encoder g and a hypothesis h. At each round, each client model is asynchronously trained and then uploaded to the server, potentially with model perturbation for protection against deep leakage attacks using differential privacy. On the server side, a client model undergoes hypothesis-transferable alignment with target data in an unsupervised manner, where the hypothesis h remains frozen while only the encoder aligns. Once a sufficient number of aligned models have been collected, a model merging process is performed and coordinated by a novel hypothesis re-weighting mechanism.

C. Client Model Training

The clients are shown in the lower part of Fig. 3. On the i-th client, a deep neural model $F_s^i : X^i \rightarrow y^i$ is training to

minimize a standard Cross-Entropy (CE) loss:

$$\mathcal{L}_{CE} = -\mathbb{E}_{(X,y)\in\mathcal{S}^i} \sum_{j\in[C]} q_j \log(\sigma_j(F_s^i(X))), \quad (1)$$

where q_j is j-th element of the true probability distribution, i.e., one-hot encoding of ground-truth label y. $\sigma_j(\cdot)$ is the j-th element of the softmax-normalized output of $F_s^i(X)$.

To improve the generalization and robustness of the model, we employ label smoothing [44] to mitigate the risk of overfitting to noisy labels. By incorporating label smoothing, the loss function in Eq. 1 can be expressed as follows:

$$\mathcal{L}_{CE} = -\mathbb{E}_{(X,y)\in\mathcal{S}^i} \sum_{j\in[C]} q_j^{lsm} \log(\sigma_j(F_s^i(X))), \quad (2)$$

where $q_j^{lsm} = (1-\eta)q_j + \eta/C$ is the smoothed label, and $\eta \leq 1$ is the smoothing factor.

D. Hypothesis Transfer

We assume that the deep neural model F follows a standard structure consisting of a feature encoding module $g : X \to \mathbb{R}^d$, where d represents the dimension of the extracted features. Additionally, it includes a classifier $h : \mathbb{R}^d \to \mathbb{R}^C$, such that the overall model can be represented as $F(X) = h(g(X))$.

1) Hypothesis Invariance of WiFi Pattern: For WiFi-based HAR, the classifier h takes charge of making HAR predictions based on the features extracted from the encoder. We argue that the classifier (even a simple linear head) is of *invariance* that is robust and generalizes well across different environments, regardless of the specific environment.

Assumption 1: **Hypothesis invariance of WiFi patterns**. Under identical hypothesis space $\mathcal{H}$, the optimal hypothesis on the source environment, i.e., $h_s^* = \arg\min_{h\in\mathcal{H}} \epsilon_{\mathcal{S}}(h)$, is also the optimal hypothesis for target environment such that $\epsilon_{\mathcal{T}}(h_s^*) \sim \inf_{h\in\mathcal{H}} \epsilon_{\mathcal{T}}(h)$.

Intuitively, the WiFi signals contain valuable information about the physical movements and activities of individuals. Despite variations in the specific signal patterns across different environments, there are certain fundamental characteristics that remain consistent and can be captured by the same classifier. For example, when the distance between the TX and RX becomes larger, it leads to weaker signal strength, but the underlying patterns that distinguish different activities remain similar. By adjusting the signal values accordingly, such as amplifying, the classifier can still recognize these patterns effectively. Similarly, when there are changes in the placement of TX and RX, there may be rotational variations in the signal, yet can be compensated for at the feature level, allowing the classifier to identify the same underlying patterns. While the specific WiFi signal patterns may vary, the invariant classifier can generalize across different environments and adapt to variations in signal strength, rotation, or other factors.

The invariant classifier assumption allows us to utilize the same classifier for both the source and target environments, i.e., $h_{inv} = h_s = h_t$. This enables us to focus on learning a domain-specific feature encoder g_s and g_t to extract critical patterns that distinguish human activities. Consequently, the model for source and target can be represented as $F_s(X) = h_{inv}(g_s(X))$ and $F_t(X) = h_{inv}(g_t(X))$, respectively.

Our model aims to democratize WiFi-based HAR for unsupervised end-user environments, formalized as a novel unsupervised federated domain adaptation problem in III-A. While potentially applicable to other domains like vision, our methodology is primarily designed for WiFi CSI data. The effectiveness of this approach outside WiFi data isn't guaranteed due to differing design focuses and requirements. WiFi-based HAR faces unique challenges, such as non-i.i.d. data and substantial domain discrepancies between source training data and target end-user environments. Our method specifically addresses these unique characteristics, targeting the non-i.i.d. problem and domain discrepancies. Therefore, we propose a new hypothesis transfer method tailored to WiFi-based HAR.

2) Aligning the Target Environment with Information Maximization Loss: Given that h_s^i is well-trained, having minimized empirical risk through training on client data, we anticipate that its outputs under the same distributional data will demonstrate two important characteristics: *intra-concentration* and *inter-distinguishability*. More specifically, for a particular instance, the output should resemble one-hot encoding, indicating a high level of certainty in its category assignment. On a larger scale, the outputs should be evenly distributed across the entire dataset, facilitating clear differentiation among different categories.

Based on the assumption that the target hypothesis h_t is invariant and equivalent to h_s^i, we expect that h_t in the target environment will demonstrate similar behavior if the source feature distribution g_s aligns well with the target distribution g_t. To achieve this alignment, we utilize the information maximization (IM) loss [45], which aims to make the target outputs intra-concentrated and inter-distinguishable.

$$\begin{aligned} \mathcal{L}_{IM} &= \mathcal{L}_{con} + \beta\mathcal{L}_{div}, \\ \mathcal{L}_{con} &= -\mathbb{E}_{X\in\mathcal{T}} \sum_{j\in[C]} \sigma_j(F_t(X)) \log(\sigma_j(F_t(X)), \\ \mathcal{L}_{div} &= KL(\hat{p}, \frac{1}{C}\mathbb{I}) - \log C, \end{aligned} \quad (3)$$

where $KL(\cdot)$ denotes the Kullback-Leibler (KL) divergence, $\mathbb{I}$ is a C-dimensional vector of ones, $\hat{p} = \mathbb{E}_{X\in\mathcal{T}}[\sigma(F_t(X)]$ denotes the mean categorical distribution over the target environment. β is a balancing factor that controls the trade-off between concentration and diversity.

It is worth noting that the client hypothesis h_s^i is uploaded to the server and remains *frozen* throughout the training process in the target environment. This can be observed in the top-left section of Fig. 3. The only module that undergoes updates is the feature encoding module g_t^i.

3) Asynchronous Alignment: Existing decentralized UMDA models, such as KD3A [16], require all client models to be ready for alignment, resulting in a time-consuming synchronous strategy. This approach is hindered by varying running speeds and the possibility of offline clients, which can significantly affect efficiency due to heterogeneous hardware and communication foundations. In contrast, our alignment process is asynchronous. Each client model is independently aligned, and the updated model is transmitted to the server

as soon as it becomes available. For example, if the client model F_s^i is on standby, we immediately align it with the target environment.

E. Hypothesis Ensemble

The alignment of hypothesis h_s^i produces the target model F_t^i. Within a given time interval, the target terminal accumulates r distinct hypothesises from clients, forming a hypothesis pool $H = \{h_s^1, h_s^2, ..., h_s^r\}$ along with their corresponding models $P = \{F_t^1, F_t^2, ..., F_t^r\}$. As the data quality and data distribution across environments may vary drastically, the optimal ensemble may not include models from all clients. We explore different strategies for selecting the ensemble.

Random Selection: The target terminal randomly selects v hypotheses from the pool and generates an ensemble by aggregating the selected models.

Source Re-weighting: This method involves a weighted sum of all the models in the pool using a weighting scheme $\boldsymbol{\alpha} \in \mathbb{R}^{1\times r}$. The ensemble can be represented as $F_t = \boldsymbol{\alpha} P$. Previous methods commonly re-weight each source domain using the $\mathcal{H}$-divergence, as shown in [46]:

$$\alpha_i = e^{-d_{\mathcal{H}}(\mathbb{D}_s^i, \mathbb{D}_t)} / \sum_{k\in[K]} e^{-d_{\mathcal{H}}(\mathbb{D}_s^k, \mathbb{D}_t)}.$$

However, as the $\mathcal{H}$-divergence needs to access source data, it contradicts our federated settings.

Hypothesis Re-weighting: Instead of using data- or gradient-focused metrics or an additional validation set [41], we propose a hypothesis-focused weighting method. Our method leverages the hypothesis invariance property, which not only holds across different environments but also remains invariant over time. At round t_0, we initialize the weights as $\boldsymbol{\alpha}^{(0)} = \frac{1}{r}\mathbb{I}$, promoting a simple averaging over the pool to yield the hypothesis $h_t^{(0)} = \boldsymbol{\alpha}^{(0)} H$ and $F_t^{(0)} = \boldsymbol{\alpha}^{(0)} P$ (as shown in the top-right part of Fig. 3). In round t_l, the weighting scheme $\boldsymbol{\alpha}$ is computed by comparing each hypothesis in the pool with the ensemble hypothesis from the previous round:

$$\alpha_i = 1 - \left[\frac{e^{|h_s^i - h_t^{(l-1)}|/\pi}}{\sum_{k\in[r]} e^{|h_s^k - h_t^{(l-1)}|/\pi}} \right], \tag{4}$$

where h_s^i represents the i-th source hypothesis from the pool, $h_t^{(l-1)}$ is the ensemble hypothesis from the previous round, is the temperature parameter, and $|\cdot|$ denotes the L2 norm. We use a dynamic temperature parameter $\pi = \frac{\pi_0}{1+\log R}$, R is the current round of communication. This helps avoid arbitrariness in the initial rounds by introducing softer weights while promoting better convergence by gradually transitioning to sharper weights as the communication continues.

F. Model Perturbation

Recent research on deep leakage [47] has highlighted risks with sharing models, particularly in terms of attacks exploiting model gradients. To address this concern, we propose a defense mechanism involving Laplacian perturbation applied to the model prior to uploading it to the server.

To ensure the security of the shared models under deep leakage, we employ the concept of differential privacy[3]. Specifically, we create a perturbation matrix having identical size $m \times n$ with the hypothesis, where each entry is independently drawn from a Laplacian distribution $Lap(\mu, b)$. Adding this random matrix to the hypothesis acts as a Johnson-Lindenstrauss transform [48], guarantees that a linear transformation of a set of n data points to a smaller subspace using a random matrix retains pairwise distances within a factor of $(1 \pm \eta)$, where η lies in the range of $[0, 0.5]$.

By introducing this perturbation, we introduce randomness and variability into the model, making it significantly more challenging for potential deep leakage attacks to extract sensitive information.

Algorithm 1: FligHT training process with round l

Input: Source environments $\{\mathbb{D}_s^k\}_{k=1}^K$, target environment $\mathbb{D}_t$, target hypothesis $h_t^{(l-1)}$
Output: Target model $F_t^{(l)}$

1 // Client Model Training
2 **foreach** $\mathbb{D}_s^k \in \{\mathbb{D}_s^k\}_{k=1}^K$ **do**
3 $h_s^k \leftarrow h_t^{(l-1)}$;
4 $g_s^k \leftarrow g_t^{(l-1)}$ or random initialization;
5 Train F_s^k with loss function $\mathcal{L}_{CE}$;
6 Add perturbation on h_s^k;
7 Upload g_s^k and h_s^k to server;
8 // Server Training
9 **foreach** h_s^k **do**
10 // Aligning the target domain
11 Initialize g_t^k by g_s^k or random initialization;
12 $F_t^k \leftarrow *h_s^k(g_t^k)$; // $*h$ indicates that it remains frozen during the training.
13 Train F_t^k with loss function $\mathcal{L}_{IM}$;
14 Compute hypothesis weight $\boldsymbol{\alpha}^{(l)}$ using Eq. 4;
15 $F_t^{(l)} \leftarrow \boldsymbol{\alpha}^{(l)}[F_t^1, F_t^2, ..., F_t^r]$;
16 **return** $F_t^{(l)}$

G. FligHT Algorithm

The integration of the components mentioned above is depicted in Algorithm 1. The process begins by obtaining the client model using the loss function $\mathcal{L}_{CE}$ (line 5). This model is derived from the updated model initialized in the previous round (lines 3-4). Subsequently, perturbations are introduced to the client model (line 6), which is then uploaded to the server (line 7).

Upon receiving each client model, the server performs asynchronous alignment with the target data using the loss function $\mathcal{L}_{IM}$ (lines 11-13). Throughout this training process, the hypothesis h_s^k remains frozen. Finally, the model $F^{(l)}$ is

[3]Notably, differential privacy is not the focus of this paper. We enhance privacy through the decentralized nature of federated learning, which ensures no direct data sharing between clients. Our model perturbation technique further safeguards privacy by concealing sensitive information in shared models.

computed by aggregating the accumulated r models, weighted according to Eq. 4 (lines 14-15).

H. Theoretical Insights

In the federated domain adaptation task, Peng et al. [15] presents a weighted error bound, denoted as $\mathcal{B}_G$, for decentralized target adaptation. It is worth noting that this bound is derived without considering our hypothesis invariance assumption. Specifically, the following theorem holds:

Theorem 1: Given hypothesis space $\mathcal{H}$ with VC-dimension d, and $\{\mathbb{D}_s^k\}_{k=1}^K$, $\mathbb{D}_t^k$ be the empirical distributions from each source environment and target environment induced by a sample of size m in a federated learning system, respectively. Then, the following bound hold for each $h \in \mathcal{H}$, any $\alpha \in \mathbb{R}^K$, $\sum_{i\in[K]} \alpha_i = 1$, with probability at least $1-\delta$ over the choice of samples:

$$\epsilon_{\mathcal{T}}(h) \leq \mathcal{B}_G$$

$$\mathcal{B}_G = \hat{\epsilon}_S(\sum_{i\in[K]} \alpha_i h_s^i) + \sum_{i\in[K]} \alpha_i (\frac{1}{2}\hat{d}_{\mathcal{H}\Delta\mathcal{H}}(\mathbb{D}_s^i, \mathbb{D}_t^i) + \lambda_i) + \tau$$

where $\hat{d}_{\mathcal{H}\Delta\mathcal{H}}(\cdot)$ measures the empirical discrepancy of distributions, $\tau = 4\sqrt{\frac{2d\log(2Km)+\log(4/\delta)}{Km}}$ is the error between empirical and generalized risks induced by the VC-dimension, $\lambda_i = \inf_{h\in\mathcal{H}}[\epsilon_{\mathcal{S}}^i(h) + \epsilon_{\mathcal{T}}(h)]$ is the error under the optimal hypothesis.

Under the hypothesis invariance assumption (Assumption 1), we establish a lower theoretical risk bound for the target environment. This leads us to the following theorem:

Theorem 2: The error bound $\mathcal{B}_{HI}$, established under Assumption 1, is a tighter bound compared to the general error bound $\mathcal{B}_G$. In other words, we have $\mathcal{B}_{HI} \leq \mathcal{B}_G$.

Proof 1: Under Assumption 1, where the source error approximates the target error under the optimal hypothesis $h^\star$, i.e., $\epsilon_{\mathcal{S}}(h^\star) \sim \epsilon_{\mathcal{T}}(h^\star)$, the bound $\mathcal{B}_{HI}$ can be updated as:

$$\begin{aligned}\mathcal{B}_{HI} =& \hat{\epsilon}_S(\sum_{i\in[K]} \alpha_i h_s^i) \\ &+ \sum_{i\in[K]} \alpha_i (\frac{1}{2}\hat{d}_{\mathcal{H}\Delta\mathcal{H}}(\mathbb{D}_s^i, \mathbb{D}_t^i) + \inf_{h\in\mathcal{H}} 2\epsilon_{\mathcal{S}}(h)) + \tau\end{aligned}$$

In accordance with Assumption 1, the risk associated with single-objective optimization is lower in comparison to the original bi-objective optimization problem. Therefore, we have the following relationship:

$$\lambda_i = \inf_{h\in\mathcal{H}}[\epsilon_{\mathcal{S}}(h) + \epsilon_{\mathcal{T}}(h)] \geq \inf_{h\in\mathcal{H}} 2\epsilon_{\mathcal{S}}(h)$$

such that $\mathcal{B}_{HI} \leq \mathcal{B}_G$ and Theorem 2 holds.

Theorem 2 shows the model has a lower risk compared to the decentralized domain adaptation models.

I. Computation and Communication Costs

Our model is computationally efficient as it only needs a one-step training to align each source-domain pair using the IM-loss described in Eq. 3. This approach eliminates additional training costs such as extra curated source and target BatchNorm MMD training in KD3A [16] and the multi-step training involving multiple encoders in FADA [15].

Furthermore, each communication between the client and server in the FligHT model incurs a cost of $\mathcal{O}(|F|)$. This cost is on par with low-cost federated domain adaptation methods such as KD3A [16] and SHOT [39], while being more efficient than FADA [15]. Unlike FADA, which requires the upload of source features along with the model, resulting in a cost of $\mathcal{O}(d\times|\mathbb{D}|+|F|)$, our FligHT model eliminates the need for such additional uploads, thereby avoiding any extra communication overhead.

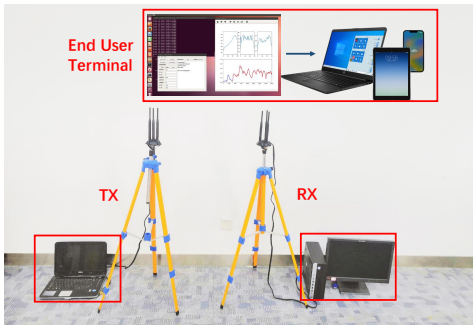


Fig. 4. The equipment of data platform. For source environments, we use a laptop and a desktop to serve as TX and RX, respectively. For the end-user, we developed an Android app that can be deployed on users' smartphones to process the computation.

IV. Experimental Validation

In this section, we evaluate the effectiveness of our model in different settings and demonstrate its superiority over state-of-the-art models. We first introduce data platform and collection settings, experimental settings, and baselines. Next, we will summarize and discuss the experimental results. Additionally, we will provide an ablative study, a feature visualization, and the categorical performance to further validate and discuss the merits of the proposed system.

A. Data Platform and Setup

The data we work with are WiFi signals, specifically WiFi Channel State Information (CSI), which varies across different environments and participants, serving as our primary data source. In the context of the IEEE 802.11n standard, CSI facilitates the transmission of information between a receiver (RX) and a transmitter (TX). A received signal y_i can be expressed as:

$$y_i = H_i x_i + n_i \tag{5}$$

where x_i is the transmitted signal, n_i is noise, and H_i is the CSI matrix for sub-carrier i:

$$H_i = \begin{pmatrix} h_i^{11} & h_i^{12} & \dots & h_i^{N_T} \\ h_i^{21} & h_i^{22} & \dots & h_i^{2N_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_i^{N_R1} & h_i^{N_R2} & \dots & h_i^{N_RN_T} \end{pmatrix} \tag{6}$$

where $h_i^{mn} = |h_i^{mn}|\, e^{j\angle h_i^{mn}}$ represents the signal measurement between the m-th RX antenna and the n-th TX antenna.

(a) Seminar Room (b) Lounge Space (c) Exhibition Space

Fig. 5. The layouts of the three environments for experiments.

1) Data Platform: For the data platform, we utilized a laptop and a desktop equipped with 5300 Network Interface Cards. The laptop served as the transmitter (TX), while the desktop acted as the receiver (RX). For CSI data collection, the RX had dedicated CSI tools installed on Ubuntu 14.04 operating system. Both the TX and RX were equipped with three antennas mounted on tripods at a height of 1.5 meters. The TX and RX were positioned 4 meters apart. To showcase the feasibility of deploying our system on edge devices, we developed an Android app for portable devices, like users' smartphones. This app receives and processes CSI data from each user, performing activity recognition locally on the smartphone. This eliminates the requirement to transmit sensitive CSI data to external servers for processing.

Fig. 4 provides an illustration of the data platform setup.

2) Experimental Environments: Our experiments were conducted in three different indoor environments: a seminar room, a lounge space, and an exhibition space. Each environment has unique spatial layouts and dimensions, as shown in Fig. 5.

- Seminar Room: The first environment was a seminar room measuring 4 meters by 3 meters, where several chairs are arranged to accommodate participants, and a meeting table is centrally positioned to facilitate discussions and collaborative activities.
- Lounge Space: The second environment was a lounge space covering an area of 10 meters by 8 meters. Within this lounge space, there is a collection of chairs and tables, creating an environment where individuals can sit, socialize, and engage in casual interactions.
- Exhibition Space: The third environment in our study was an exhibition space, designed to accommodate a wide range of exhibits and displays. This space consisted of a central area measuring 5.5 meters by 18.5 meters and an additional 8-meter extension. The surrounding walls were adorned with displays, and there were also dedicated areas for additional displays within the exhibition space.

3) Data Acquisition: In our experiments, the router was equipped with three antennas, each having 30 channels. A timestamp reading had a dimension of 1×90. To capture CSI records for different activities, we employed a sliding window approach with a window size of 4 seconds. This allowed us to gather comprehensive data. With a data collection frequency set at 500 Hz, each sample had a shape of 2000×90. A total of four volunteers participated in our experiment, consisting of two males and two females. They were graduate/undergraduate students aged between 18 and 27, with heights ranging from 1.6 meters to 1.8 meters. The experiment involved seven common daily activities: "Standing still", "Jump", "Bend", "Run", "Sit down", "Walk", and "Wave hand". Each volunteer is asked to perform each activity 101 times in each environment. The data collection process spanned six weeks, with a total of 4949 samples collected for each environment and a combined total of 14847 samples.

4) CSLOS dataset: *CSLOS* dataset [49] includes data from three indoor environments with 30 participants. The WiFi devices were Intel 5300 network cards, transmitting signals via one antenna and capturing them with three antennas. Two environments were in a Line-of-Sight (LOS) setting, while the third was in a Non-Line-of-Sight (NLOS) setting. The collected data included RSSI and CSI values from Wi-Fi signals exchanged during 5 experiment trials conducted by each of the 10 participants per environment.

5) Additional Sensory Dataset: To validate the generality for different types of data sources, we also included an additional sensory dataset *UCIHAR* [50]. The *UCIHAR* dataset comprises sensor readings collected from smartphones worn by 30 participants aged between 19 and 48 years. The experiment involved six activities, consisting of three static postures (standing, sitting, and lying) and three dynamic activities (walking, walking downstairs, and walking upstairs). Participants wore smartphones on their waist, which recorded 3-axial linear acceleration, 3-axial angular velocity, and readings from an embedded accelerometer and gyroscope at a sampling rate of 50Hz. n this scenario, as different individuals may exhibit unique characteristics and subtle nuances in their activities, leading to variations in the underlying data distribution among subjects, we treated each person as a separate domain.

B. Implementation Details and Evaluation Metrics

We adopted the standard FL protocol to extend the centralized method to our federated setting. Our feature encoder $g(\cdot)$ consists of two convolutional layers with 16 and 32 kernels of size 5×5 respectively. Batch normalization, dropout with a rate of 0.5, ReLU activation, and max pooling with a kernel size of 2, were applied after each convolutional layer. Additionally, we included an extra linear layer with a hidden dimension

TABLE I
MAIN RESULT: QUANTITATIVE COMPARISON WITH BASELINES ON ***Cross-Person*** SETTING. WE TAKE TURNS USING ONE PERSON'S DATA AS THE TARGET AND THE REST PERSONS' AS SOURCES AND REPORT THE AVERAGE ACCURACY (%) ON THE TARGET. REMARKABLY, BOTH TWO VARIANTS OF OUR MODEL (FLIGHT, AND FLIGHT-MP), SURPASS BOTH THE CENTRALIZED AND FEDERATED UPPER BOUNDS, ACHIEVING SIGNIFICANTLY HIGHER ACCURACY THAN ALL THE BASELINES.

Type	Methods	$P_1P_2P_3 \rightarrow P_4$	$P_1P_2P_4 \rightarrow P_3$	$P_1P_3P_4 \rightarrow P_2$	$P_2P_3P_4 \rightarrow P_1$	Average
Centralized	Source Only	70.06	79.59	74.82	63.26	71.93
	DANN [27]	79.59	82.99	75.51	68.02	76.53
	SHOT [40]	89.79	<u>92.51</u>	90.47	83.67	89.11
Federated	FedAvg [10]	59.86	78.91	64.62	55.78	64.79
	FedCHAR-DC [35]	60.90	60.90	61.61	58.39	60.45
	F-DANN [27]	68.02	76.19	71.42	59.86	68.87
	FADA [15]	64.97	78.91	82.43	68.75	73.77
	KD3A [16]	72.80	86.40	88.29	73.50	80.25
	FedProto [33]	76.20	81.29	89.80	81.00	82.07
	FligHT	**89.12**	**94.50**	**95.43**	**87.76**	**91.70**
	FligHT-MP	<u>88.44</u>	91.10	<u>94.43</u>	<u>84.35</u>	<u>89.58</u>

TABLE II
MAIN RESULT: QUANTITATIVE COMPARISON WITH BASELINES ON ***Cross-Environment*** SETTING. WE TAKE TURNS USING ONE INDOOR ENVIRONMENT AS THE TARGET AND THE REST ENVIRONMENT AS SOURCES AND REPORT THE AVERAGE ACCURACY. OUR MODELS (FLIGHT, AND FLIGHT-MP) SURPASS BOTH THE CENTRALIZED AND FEDERATED UPPER BOUNDS, ACHIEVING SIGNIFICANTLY HIGHER ACCURACY THAN ALL THE BASELINES.

Type	Methods	$E_1E_2 \rightarrow E_3$	$E_1E_3 \rightarrow E_2$	$E_2E_3 \rightarrow E_1$	Average
Centralized	Source Only	36.90	44.11	36.96	39.32
	DANN [27]	47.68	<u>78.78</u>	76.02	67.49
	SHOT [40]	55.07	**80.13**	<u>83.57</u>	<u>72.92</u>
Federated	FedAvg [10]	34.68	38.59	35.55	36.27
	FedCHAR-DC [35]	36.76	36.33	36.63	36.57
	F-DANN [27]	44.21	42.17	48.29	44.89
	FADA [15]	35.46	61.48	41.30	46.08
	KD3A [16]	51.00	64.00	58.50	57.83
	FedProto [33]	**64.95**	68.42	59.76	64.38
	FligHT	<u>59.73</u>	78.18	**85.52**	**74.48**
	FligHT-MP	52.53	75.42	76.23	68.06

TABLE III
MAIN RESULT: QUANTITATIVE COMPARISON WITH BASELINES ON ***Cross-Environment*** SETTING IN OPEN-SOURCE DATASET CSLOS [49]. OUR MODELS (FLIGHT, AND FLIGHT-MP) SURPASS BOTH THE CENTRALIZED AND FEDERATED UPPER BOUNDS, ACHIEVING SIGNIFICANTLY HIGHER ACCURACY (%) THAN ALL THE BASELINES.

Type	Methods	$E_1E_2 \rightarrow E_3$	$E_1E_3 \rightarrow E_2$	$E_2E_3 \rightarrow E_1$	Average
Centralized	Source Only	36.25	25.00	37.50	32.92
Federated	F-DANN [27]	39.17	38.33	47.50	41.67
	FADA [15]	49.10	37.66	59.00	48.59
	KD3A [16]	53.30	41.70	61.70	52.23
	FedProto [33]	61.17	62.16	63.67	62.33
	FligHT	70.00	58.33	75.00	67.78
	FligHT-MP	65.00	58.10	73.33	65.48

TABLE IV
ABALTION STUDY: AVERAGE ACCURACY (%) OF USING DIFFERENT WEIGHTING STRATEGIES IN FLIGHT MODEL.

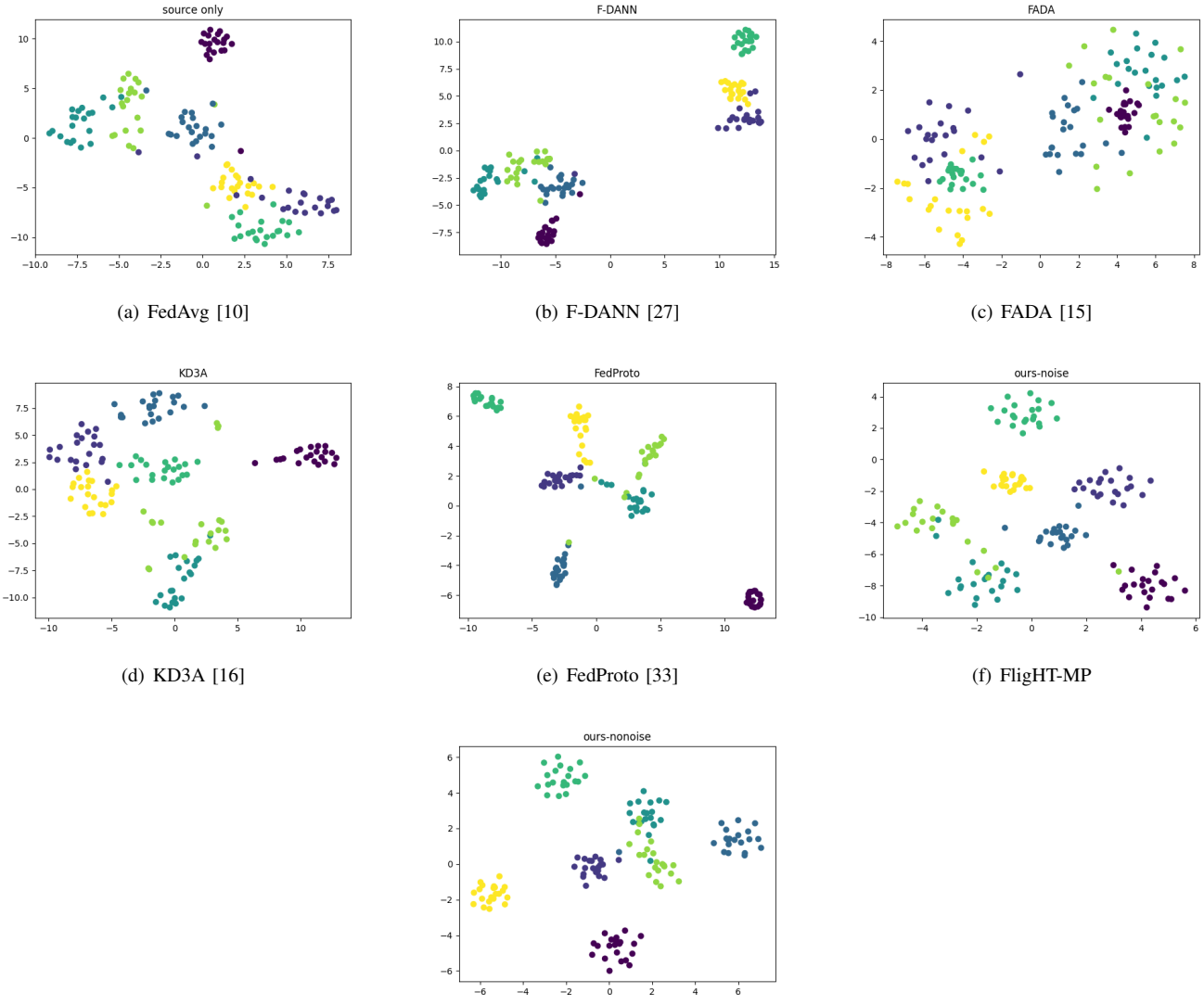
Method	Cross-Person		Cross-Environment	
	ACC	Δ	ACC	Δ
Hypothesis Re-weighting	**91.45**	-	**74.81**	-
Averaging	90.01	1.44	72.86	1.95
Random Selection	87.68	3.77	69.77	5.04

of 256. The classifier, denoted as $h(\cdot)$, is a linear head used to project the features into the corresponding categories. Our model complexity is comparable to the decentralized UMDA baselines SHOT and KD3A, while FADA has a significantly larger model complexity due to its additional disentangle module.

We follow the common practice in federated learning [10] to simulate the clients: Each client trains the local model using a mutually exclusive data setting (i.e., an environment or a person), resulting in non-i.i.d. data partitioning. The clients are trained in parallel as described in Section III-C, using dynamic client selection and the aggregation strategy outlined in Section III-E.

We utilized the Adam optimizer with an initial learning rate of 0.001. The label smoothing factor $\eta = 0.1$. The default training epoch was set to 50 for clients and 5 for the server.

In terms of model perturbation, we set the parameters of the Laplacian distribution with $\mu = 0$ and $b = 0.01$ for the



(a) FedAvg [10] (b) F-DANN [27] (c) FADA [15]

(d) KD3A [16] (e) FedProto [33] (f) FligHT-MP

(g) FligHT

Fig. 6. **Feature Visualization**: t-SNE plot of features of (a) FedAvg, (b) F-DANN, (c) FADA, (d) KD3A, (e) FedProto, (f) FligHT-MP, and (g) FligHT in cross-person $P_1P_3P_4 \rightarrow P_2$ setting. We use different colors to denote different categories.

cross-person setting, and $b = 0.001$ for the cross-environment setting. The loss balance factor was $\beta = 1$, and the initial temperature was $\pi_0 = 2$.

The dataset includes four individuals performing seven activities in three environments, with each activity repeated 101 times per individual per environment. This results in 4,949 samples per environment and a total of 14,847 samples. For training and testing, we use a source-target adaptation setting, rotating one person/environment as the target and the rest as sources. We report recognition accuracy on the target environment, i.e., the proportion of correctly identified activities among all predictions.

C. Baselines

We compared our model with eight state-of-the-art (SOTA) baselines, categorized into centralized and federated approaches. In the centralized category, the model is trained using all client data aggregated in a central server. This category contains three SOTA models: centralized Source Only, directly apply the client model to the target environment without further adaptation; adversarial domain adaptation model DANN [27]. We also include the SOTA hypothesis-based method SHOT [40] from the ML community for comparison. SHOT operates under a centralized learning paradigm, generating a single hypothesis for adaptation. In contrast, our model follows a multi-source federated structure, generating multiple hypotheses that are synchronized with the target data using innovative hypothesis ensemble methods.

In the federated category, the data is decentralized and kept on each client, we considered the following models: Federated Averaging (FedAvg) [10], which is the federated version of source-only that does not have additional adaptation; FedCHAR-DC [35], a federated clustering model that group clients having similar data distribution into clusters for

TABLE V
CATEGORICAL PERFORMANCE: CONFUSION MATRICES OF THE PROPOSED FLIGHT MODEL AND BASELINES ON ***Cross-Person*** SETTING ($P_1P_3P_4 \rightarrow P_2$).

(a) FedAvg [10]

Ground-truth \ Predicted	Standing still	Jump	Bend	Run	Sit down	Walk	Wave hand
Standing still	0.95	0.00	0.00	0.00	0.05	0.00	0.00
Jump	0.00	0.52	0.00	0.10	0.00	0.00	0.38
Bend	0.05	0.19	0.10	0.05	0.00	0.47	0.14
Run	0.00	0.00	0.00	0.81	0.00	0.19	0.00
Sit down	0.00	0.57	0.00	0.00	0.33	0.00	0.10
Walk	0.05	0.00	0.00	0.09	0.00	0.86	0.00
Wave hand	0.00	0.00	0.00	0.00	0.00	0.05	0.95

(b) F-DANN [27]

Ground-truth \ Predicted	Standing still	Jump	Bend	Run	Sit down	Walk	Wave hand
Standing still	0.76	0.00	0.05	0.00	0.00	0.19	0.00
Jump	0.00	0.48	0.05	0.00	0.00	0.05	0.58
Bend	0.00	0.00	0.71	0.05	0.00	0.05	0.19
Run	0.00	0.05	0.00	0.67	0.00	0.28	0.00
Sit down	0.00	0.00	0.00	0.00	0.67	0.10	0.23
Walk	0.00	0.00	0.00	0.00	0.00	0.76	0.24
Wave hand	0.00	0.00	0.00	0.00	0.00	0.05	0.95

(c) FADA [15]

Ground-truth \ Predicted	Standing still	Jump	Bend	Run	Sit down	Walk	Wave hand
Standing still	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Jump	0.00	0.89	0.00	0.00	0.11	0.00	0.00
Bend	0.11	0.00	0.74	0.00	0.00	0.15	0.00
Run	0.00	0.00	0.00	0.89	0.00	0.11	0.00
Sit down	0.00	0.16	0.00	0.00	0.84	0.00	0.00
Walk	0.00	0.00	0.00	0.02	0.05	0.88	0.05
Wave hand	0.00	0.18	0.00	0.00	0.29	0.00	0.53

(d) KD3A [16]

Ground-truth \ Predicted	Standing still	Jump	Bend	Run	Sit down	Walk	Wave hand
Standing still	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Jump	0.00	0.90	0.05	0.00	0.00	0.00	0.05
Bend	0.05	0.00	0.95	0.00	0.00	0.00	0.00
Run	0.00	0.00	0.00	0.62	0.00	0.38	0.00
Sit down	0.00	0.00	0.00	0.00	0.95	0.00	0.05
Walk	0.14	0.00	0.05	0.00	0.05	0.76	0.00
Wave hand	0.00	0.00	0.00	0.00	0.00	0.00	1.00

(e) FedProto [33]

Ground-truth \ Predicted	Standing still	Jump	Bend	Run	Sit down	Walk	Wave hand
Standing still	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Jump	0.00	0.86	0.00	0.00	0.00	0.14	0.00
Bend	0.00	0.00	0.95	0.00	0.00	0.05	0.00
Run	0.00	0.05	0.00	0.81	0.10	0.05	0.00
Sit down	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Walk	0.00	0.00	0.00	0.24	0.05	0.67	0.05
Wave hand	0.00	0.00	0.00	0.00	0.00	0.00	1.00

(f) FligHT

Ground-truth \ Predicted	Standing still	Jump	Bend	Run	Sit down	Walk	Wave hand
Standing still	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Jump	0.00	0.97	0.02	0.00	0.00	0.00	0.00
Bend	0.00	0.00	0.96	0.00	0.00	0.03	0.00
Run	0.00	0.00	0.00	0.91	0.00	0.10	0.00
Sit down	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Walk	0.00	0.00	0.00	0.16	0.00	0.84	0.00
Wave hand	0.00	0.00	0.00	0.00	0.00	0.00	1.00

(g) FligHT-MP

Ground-truth \ Predicted	Standing still	Jump	Bend	Run	Sit down	Walk	Wave hand
Standing still	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Jump	0.00	0.95	0.00	0.05	0.00	0.00	0.00
Bend	0.00	0.00	0.95	0.05	0.00	0.00	0.00
Run	0.00	0.00	0.00	0.90	0.00	0.10	0.00
Sit down	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Walk	0.00	0.00	0.00	0.10	0.05	0.81	0.05
Wave hand	0.00	0.00	0.00	0.00	0.00	0.00	1.00

mitigating the non-i.i.d. problem; F-DANN [27], which is the federated version of DANN; federated adversarial domain adaptation mode FADA [15]; a knowledge distillation-based domain adaptation model KD3A [16]; and a heterogeneous federated learning model FedProto [33].

It is worth noting that F-DANN, FADA, and FedProto are considered quasi-federated models. These models have a requirement for sharing source features (F-DANN and FADA) or prototypes[4] (FedProto) between the source and target domains. In contrast, models such as KD3A and our own model, FligHT, do not permit feature sharing. This distinction sets KD3A and FligHT apart from the quasi-federated models, as they adopt different approaches that do not involve feature sharing between domains.

[4]Prototypes are 1D-vectors generated by averaging the low-dimension representations of samples from the same class.

D. Main Results

We conduct experiments to verify the effectiveness of our model on both *Cross-Person* and *Cross-Environment* settings, which take turns using one person/environment as the target and the rest person/environment as sources and report the average accuracy. We compare two versions of our FligHT model *with* or *without* model perturbation, denoted as FligHT-MP, and FligHT, respectively.

1) Cross-Person: The results for the *Cross-Person* setting are presented in Table I. We can see that both variants of our model demonstrate remarkable performance surpassing all baselines and achieving the highest scores in this setting. Within the federated paradigm, FligHT and FligHT-MP outperform the previously best model, FedProto, by an average margin of 9.63 and 7.51 percentage points, respectively. Furthermore, when compared to the top-performing centralized model, SHOT, our models FligHT and FligHT-MP still exhibit superior performance, achieving margins of 2.59 and 0.47 respectively. Compared with federated clustering model FedCHAR-DC [35], our models (i.e., FligHT and FligHT-MP) still achieved a significantly higher accuracy with a 30% above performance gap. The substantial performance gap can be attributed to the severe domain shifts problem in WiFi-based HAR, which is not explicitly addressed in the federated learning paradigm. These impressive results significantly advance the state-of-the-art accuracy, increasing it from 82.07% to 91.70%. Consequently, these findings strongly support the practical application of Human Activity Recognition (HAR) in real-world end-user environments, emphasizing the anticipated superior performance of WiFi-based HAR.

When comparing the federated and centralized paradigms, it is expected that federated models would generally underperform their centralized counterparts, as observed in the cases of FedAvg v.s. Source Only and F-DANN v.s. DANN. This aligns with the *FedAvg Theorem*, stating that centralized models serve as an upper bound for federated models. Notably, when confronted with more significant domain shifts, such as $P_1P_2P_4 \rightarrow P_3$ and $P_1P_3P_4 \rightarrow P_2$ settings, the performance gaps between the two paradigms become more pronounced. This suggests domain shifts on WiFi data have a negative effect on model performance and highlights the importance of domain alignment. In contrast, our model exhibits significantly better performance in these challenging settings, achieving accuracies of 94.50% and 95.43% respectively. This demonstrates the effectiveness of the IM loss and the utilization of hypothesis transfer, highlighting our model's capability to mitigate domain shifts and enhance overall performance.

2) Cross-Environment: Table II presents results on the *Cross-Environment* setting. The results show our models continue to outperform all baselines in this setting as well. Under the federated paradigm, FligHT and FligHT-MP surpass the best-performing baseline FedProto by average margins of 10.1 and 3.68 percentage points, respectively. Moreover, when compared to the top-performing decentralized domain adaptation model, KD3A, our model demonstrates superiority with margins of 16.65 and 10.23 percentage points respectively. Additionally, in comparison to the best centralized model, SHOT, FligHT achieves an average margin of 1.56 percentage points. These results highlight the effectiveness of our models across various cross-settings, emphasizing their ability to consistently outperform competing approaches.

Another observation is, cross-environment scenarios often involve larger domain shifts, leading to relatively lower recognition accuracies for all baseline models, including the centralized ones, which fall below 75%. Notably, the $E_1E_2 \rightarrow E_3$ setting exhibits the poorest overall performance, with accuracies falling below 65%. This should ascribe to the distinctive physical characteristics of the third environment, i.e., E_3, which significantly differs from the first two environments (an empty and semi-enclosed exhibition space versus small and enclosed room spaces, as illustrated in Fig. 5). Consequently, this significant domain shift poses a considerable challenge. This observation further explains why the dedicated heterogeneous federated learning model, FedProto, achieves the best performance in this specific setting. However, our model still achieves the second-best performance, demonstrating its effectiveness in addressing significant domain shifts and the contribution of our proposed approaches.

3) CSLOS Dataset: The results on open-source dataset CSLOS [49] are detailed in Table III. The results show a consistent trend across both self-collected and open-source datasets. Remarkably, both two variants of our model (FligHT, and FligHT-MP), outperform the best-performing baselines in both centralized and federated settings, achieving significantly higher accuracy than all the baselines. but also achieve significantly higher accuracy compared to all baseline models.

E. Ablation Studies

The ablation study for different weighting strategies in the FligHT model is presented in Table IV. Among the tested strategies, the default hypothesis re-weighting strategy achieves the highest accuracy, with 91.45% in the cross-person setting and 74.81% in the cross-environment setting. Comparatively, the simple averaging strategy shows a slight decrease of 1.44 and 1.95 percentage points in the two cross-settings respectively. Due to its inherent arbitrariness, the random selection strategy, which randomly selects two hypotheses for cross-person and one hypothesis for cross-environment for the hypothesis ensemble, exhibits a significant margin and unstable results when compared to our hypothesis re-weighting approach. This random selection strategy results in a decrease of 3.77 and 5.04 points in the respective settings. Overall, the hypothesis re-weighting strategy consistently outperforms the other weighting strategies, demonstrating its effectiveness in improving accuracy and reinforcing its significance in the FligHT model.

The ablation study for model perturbation can be observed by comparing the performance of FligHT-MP (with model perturbation) and FligHT (without model perturbation). The corresponding results are listed in the last two rows of Table I and Table II. It suggests the addition of model perturbation leads to a decline in recognition performance. Specifically, there is a decrease of 2.63 points and 6.75 points in the Cross-Person and Cross-Environment settings, respectively.

The larger decline in the Cross-Environment setting is attributed to the greater domain shift and the perturbation making the inference of informative patterns challenging. It is important to note that there is no need for a direct comparison with baselines as we are the first to consider deep leakage attacks in the decentralized UMDA scenario. Despite this, the perturbated model FligHT-MP still outperforms the best baselines by a significant margin of 13.26 points in the Cross-Person setting and 7.57 points in the Cross-Environment setting.

F. Feature Visualization

We visualize t-SNE feature embeddings in Fig. 6 to analyze the effectiveness of representation learning in our model compared to other federated baselines. Fig. 6 plots the feature embeddings on target data in the cross-person $P_1P_3P_4 \rightarrow P_2$ setting, including FedAvg, F-DANN, FADA, KD3A, FedProto, FligHT-MP, and FligHT models, with each category indicated by a different color.

From Fig. 6, we can observe that the plots of state-of-the-art UMDA models, such as FADA and KD3A, exhibit blurred boundaries with large intra-class variances. This can be attributed to the fact that they solely rely on marginal alignment, which prevents them from achieving class-level separation. For federated learning models, such as FedProto and F-DANN, although they demonstrate smaller intra-class variances, some classes appear to be closely clustered, resulting in reduced discriminability. In contrast, both variants of our model exhibit more distinct decision boundaries, characterized by smaller intra-class variances and larger inter-class variances. This can be attributed to the effectiveness of our Information Maximization (IM) loss, which considers both intra-concentration and inter-distinguishability. The results clearly demonstrate that our model's design successfully generates the desired properties of feature embeddings and effectively extracts informative features in the target environment. This accounts for its superiority compared to other baselines.

G. Categorical Performance

To ensure a detailed and balanced performance analysis across different categories, we conducted a categorical comparison to complement the main results in TableI. While TableI shows only average performance, our categorical comparison in Table V compares the proposed FligHT model and baselines for each category. This comparison is essential for a comprehensive assessment, revealing strengths and weaknesses across different activities and preventing high overall accuracy from being skewed by the majority class. Additionally, it helps in detecting biases, crucial for fairness and equal performance across categories.

Table V shows the confusion matrices of our models (FligHT and FligHT-MP) compared to several baselines. The results demonstrate that both variants of our model outperform all baselines in accurately identifying activities. In addition, our model exhibits consistently good performance in distinguishing between different activities, with accuracies consistently above 80%. This suggests a consistent and robust categorical performance of our model in capturing and aligning conditional patterns associated with each activity. In other words, our models do not solely rely on specific biases but instead demonstrate the ability to capture the general patterns of each activity.

Another observation is that state-of-the-art federated domain adaptation models suffer from performance degradation in certain activities. For instance, the KD3A model encounters difficulties in accurately recognizing similar activities such as running and walking. This can ascribe to the direct alignment of target models in KD3A, which can lead to negative transfer and canceling effects. Similarly, the FADA model tends to misclassify waving hands as sitting down. This may result from an overemphasis on disentangling domain- and task-invariant features, which may inadvertently lead to information loss, as a consequence, further worsen the already weak and subtle patterns in WIFI signals, particularly for nuanced activities such as waving hands.

TABLE VI
CLIENT EFFICIENCY: COMPARISON OF AVERAGE ACCURACY (%) FOR FADA, KD3A, AND FLIGHT WITH VARYING NUMBERS OF CLIENTS.

Method	Number of Clients			Average
	2	4	6	
FADA [15]	60.94	76.56	77.14	71.53
KD3A [16]	70.70	81.00	88.40	80.03
FedProto [33]	**81.63**	88.44	90.48	**86.85**
FligHT	68.70	**90.48**	**95.92**	85.02

H. Client Efficiency

We conduct a client efficiency analysis to verify performance under different numbers of clients. The numbers of clients are set to $\{2$ $(P_1P_3 \rightarrow P_2)$, 4 $(P_1P_3P_4P_5 \rightarrow P_2)$, 6 $(P_1P_3P_4P_5P_6P_7 \rightarrow P_2)\}$. We compare with two UMDA baselines FADA and KD3A, and one federated learning baseline FedProto. The results are listed in Table VI.

From the results, we can observe that when the number of clients is extremely limited, such as with only 2 clients, the client efficiency of all UMDA models is not competitive compared to the federated learning model FedProto. This can be attributed to the data augmentation effect provided by the prototype sharing mechanism in FedProto, which helps reduce the risk of overfitting in such limited client scenarios. Nevertheless, as we increase the number of clients to 4, our model shows better client efficiency over the others. Specifically, our model excels FedProto by a margin of 2.04 points and KD3A by 9.48 points. Furthermore, when utilizing all 6 clients, the advantages of our model become even more apparent, with the margin between FedProto enlarging to 5.44. Overall, with a moderate number of clients, such as 4 and 6, the UMDA models prove to be effective in mitigating domain shifts and outperform FedProto. Our model demonstrates generally good client efficiency compared with state-of-the-art baselines.

I. Convergence Analysis

Table VIII provides a convergence analysis to show communication rounds required to reach four target accuracy levels

TABLE VII
QUANTITATIVE COMPARISON WITH BASELINES ON *Cross-Person* SETTING, USING SENSORY SIGNAL FROM *UCIHAR* DATASET.

Type	Methods	$P_1P_2P_3 \rightarrow P_4$	$P_1P_2P_4 \rightarrow P_3$	$P_1P_3P_4 \rightarrow P_2$	$P_2P_3P_4 \rightarrow P_1$	Average
Centralized	Source Only	59.34	87.37	75.82	72.38	73.73
	DANN [27]	64.58	63.10	76.92	83.80	72.10
	SHOT [40]	<u>88.54</u>	92.71	94.32	88.57	91.04
Federated	FedAvg [10]	50.67	55.88	35.07	54.13	48.94
	F-DANN [27]	52.94	65.96	67.29	68.59	63.70
	FADA [15]	81.05	84.01	90.03	75.24	82.58
	KD3A [16]	**89.30**	90.30	91.22	82.90	88.14
	FedProto [33]	85.39	89.15	91.80	91.91	89.56
	FligHT	87.50	**96.88**	**96.59**	**96.15**	**94.28**
	FligHT-MP	84.38	<u>93.75</u>	<u>95.45</u>	<u>94.23</u>	<u>91.95</u>

TABLE VIII
CONVERGENCE ANALYSIS: THE NUMBER OF COMMUNICATION ROUNDS REQUIRED TO REACH A CERTAIN ACCURACY (%). TESTED ON CROSS-PERSON $P_1P_3P_4 \rightarrow P_2$ SETTING. AN ACCURACY LEVEL IS DENOTED AS $\emptyset$ TO INDICATE THAT IT IS SKIPPED, WHILE $\neg$ IS USED TO INDICATE THAT IT IS NOT REACHABLE FOR CERTAIN METHODS.

Method	Target Accuracy			
	50%	70%	80%	90%
FADA [15]	$\emptyset$	3	13	$\neg$
KD3A [16]	6	15	19	$\neg$
FedProto [33]	3	5	7	$\neg$
FligHT	6	$\emptyset$	$\emptyset$	8

in $\{50\%, 70\%, 80\%, \text{and } 90\%\}$. In this analysis, an accuracy level is represented as $\emptyset$ when it is skipped or not considered, and $\neg$ when it is not reachable for certain methods.

From Table VIII, we can see that the FADA method skipped the 50% accuracy level altogether and directly achieved a 70% accuracy level in just 3 communication rounds. However, reaching the 80% accuracy level required 13 communication rounds, indicating a slower convergence. Unfortunately, the 90% accuracy level was not reachable using the FADA method. The FedProto shows a good communication efficiency for the first few round, as it only requires 3 rounds to reach the 50% accuracy level, and takes 5 and 7 rounds to reach the 70% and 80% accuracy levels. Despite this, the model is unable to achieve the 90% accuracy level. For the KD3A and our FligHT models, both methods required 6 communication rounds to reach the 50% accuracy level. However, the KD3A method took 15 rounds to achieve a 70% accuracy level, and 19 rounds to reach 80%. In contrast, our method surpassed these intermediate levels and directly reached the 90% accuracy level, which was not achievable for any of the baseline methods.

This indicates our model's better convergence behavior in terms of the number of communication rounds required to achieve significantly higher accuracy levels. Our approach demonstrates the ability to bypass certain accuracy thresholds and reach higher levels of accuracy with fewer communication rounds, suggesting its superiority in terms of convergence speed and efficiency.

J. Additional Experimental on Sensory Dataset

To comprehensively evaluate the effectiveness and robustness of our proposed model across different data sources, we conducted additional experiments on the widely used *UCI-HAR* datasets, as shown in Table VII. The results demonstrate a similar trend to experiments on WiFi data, with both variants of our model consistently outperforming all baselines. Specifically, under the federated paradigm, FligHT and FligHT-MP achieve an average performance improvement of 4.72 and 2.39 percentage points, respectively, compared to the best previous model FedProto. Moreover, when compared to the best model SHOT under the centralized paradigm, our model has performance gains of 3.24 and 0.91 percentage points, respectively. The above observations clearly indicate that our model not only shows high effectiveness on WiFi data but also highlights good generalizability on different data sources.

In addition, we simulated all 30 persons, with a designating one as the target person (e.g., P_1) and using the remaining 29 persons as sources. The results are listed in Table IX. The results demonstrate a consistent trend with our experiments on WiFi data and 4 client settings, showing our model outperforms all baselines. In the centralized setting, the Source Only model achieves an average accuracy of 81.81%. Under the federated paradigm, F-DANN and FedProto achieve average accuracies of 89.85% and 91.63%, respectively. Our proposed models, FligHT, significantly outperform the baselines, achieving an average accuracy of 97.41%. Specifically, FligHT shows improvements in individual accuracy for P_1, P_2, P_3, and P_4, with accuracies of 98.16%, 97.77%, 98.90%, and 94.79%, respectively. Among federated methods, FligHT achieves an average accuracy improvement of 5.78 percentage points over FedProto and 7.56 percentage points over F-DANN. Compared with the centralized method, FligHT demonstrates a substantial improvement over the Source Only model in the centralized setting by 15.60 percentage points. These observations indicate that our model not only demonstrates high effectiveness on WiFi data but also exhibits good generalizability across different data sources, as its superior performance on the *UCIHAR* dataset.

K. Communication and Computation Cost

We verified the communication cost of each interaction between the client and server, measured by the transferred model size. The results are listed in Table X. The results align with the formal computational analysis in Section III-I. Each communication between the client and server in the FligHT model incurs a cost equivalent to the model size without any

TABLE IX
QUANTITATIVE COMPARISON WITH BASELINES ON *Cross-Person* SETTING, USING SENSORY SIGNAL FROM *UCIHAR* DATASET. WE USE ALL 30 PERSONS, DESIGNATING ONE AS THE TARGET PERSON (E.G., P_1) AND USING THE REMAINING 29 PERSONS AS CLIENTS.

Type	Methods	P_1	P_2	P_3	P_4	Average
Centralized	Source Only	84.76	83.52	91.26	67.71	81.81
Federated	F-DANN [27]	95.24	83.52	94.17	86.46	89.85
	FedProto [33]	89.60	93.80	92.50	90.60	91.63
	FligHT	98.16	97.77	98.90	94.79	97.41

TABLE X
COMMUNICATION COST PER CLIENT-SERVER INTERACTION, MEASURED BY THE TRANSFERRED MODEL SIZE.

	FADA [15]	KD3A [16]	FedProto [33]	FligHT
Commu. Cost	240.33 M	153.83 M	32.79 M	33.57 M

additional communication cost. This cost is comparable to lightweight federated domain adaptation methods like FedProto (approximately 33 M), while being more efficient than FADA, which requires transferring source features in addition to the model (240.33 M vs. 33.57 M, 7.16 times smaller). Regarding KD3A, it inherently has a larger model size than ours, resulting in a cost of 153.83 M. This indicates that our FligHT model eliminates the need for additional uploads, thereby avoiding any extra communication overhead.

We measured the training time (in seconds) required to achieve a certain accuracy to evaluate the computational cost. The results are listed in Table XI. From these results, we can see that our model is significantly faster per training round compared to the two state-of-the-art federated domain adaptation (Federated DA) methods, KD3A and FADA. This is because our model does not need to compute source features (FADA) or synthesize sources (KD3A) during training, and it performs comparably with the federated-only model FedProto, which does not require an adaptation process. The difference in running time becomes more obvious when considering the final performance. Our model achieves 90% accuracy in just 66.32 seconds, whereas FADA and KD3A take over 200 seconds to merely reach 80% accuracy, and 90% accuracy is also unreachable for FedProto. This demonstrates that our model has high computational efficiency compared to state-of-the-art models.

V. CONCLUSION

We propose FligHT, a federated WiFi-based HAR model, designed to improve performance in the target environment while ensuring data privacy and efficient resource utilization. FligHT leverages the hypothesis invariant assumption of WiFi data and incorporates asynchronous hypothesis transfer, lightweight hypothesis ensemble, and model perturbation to address negative transfer and protect against deep leakage attacks. Our model achieves a tighter theoretical upper bound compared to existing multi-source federated domain adaptation models. Extensive experiments conducted on a realistic end-user emulation platform demonstrate substantial improvements over state-of-the-art baselines, with performance gains of approximately 10 percentage points in cross-person and cross-environment scenarios.

Although the model is designed to be lightweight, it still demands computational resources that might be challenging for extremely resource-constrained edge devices. Another issue is real-time adaptation, which needs improvement, as the model may experience latency when adjusting to dynamic changes. Future research should focus on proposing advanced domain adaptation techniques to further optimize our model for resource-constrained devices and enhance real-time adaptation capabilities.

VI. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grant No. 62476053; Key Program for International Cooperation of Ministry of Science and Technology, China under Grant No. 2024YFE0100700. Dr. Bing Li was with CFAR, A*STAR, Singapore, and IHPC, A*STAR, Singapore. Dr. Bing Li is now with the UESTC.

REFERENCES

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE communications surveys & tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.
[2] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey of human activity recognition using wifi csi," *arXiv preprint arXiv:1708.07129*, 2017.
[3] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.
[4] B. Liu, H. Cai, Z. Ju, and H. Liu, "Rgb-d sensing based human action and interaction analysis: A survey," *Pattern Recognition*, vol. 94, pp. 1–12, 2019.
[5] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: Human identification using wifi signal," in *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2016, pp. 75–82.
[6] S. K. Yadav, S. Sai, A. Gundewar, H. Rathore, K. Tiwari, H. M. Pandey, and M. Mathur, "Csitime: Privacy-preserving human activity recognition using wifi channel state information," *Neural Networks*, vol. 146, pp. 11–21, 2022.
[7] L. Zhang, W. Cui, B. Li, Z. Chen, M. Wu, and T. S. Gee, "Privacy-preserving cross-environment human activity recognition," *IEEE Transactions on Cybernetics*, 2021.
[8] Q. Shen, H. Feng, R. Song, S. Teso, F. Giunchiglia, H. Xu *et al.*, "Federated multi-task attention for cross-individual human activity recognition," in *IJCAI*. IJCAI, 2022, pp. 3423–3429.
[9] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 2031–2063, 2020.
[10] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

TABLE XI
TRAINING TIME COST (IN SECONDS) FOR ACHIEVING CERTAIN ACCURACY. TESTED ON OUR SELF-COLLECTED DATASETS UNDER CROSS-PERSON $P_1P_3P_4 \rightarrow P_2$ SETTING. $\emptyset$ INDICATES THAT IT IS SKIPPED, AND $\neg$ INDICATES THAT IT IS NOT REACHABLE FOR CERTAIN METHODS.

Type	Method	Training Time/Round	Time Cost for Achieving Certain Accuracy			
			50%	70%	80%	90%
Federated only	FedProto [33]	6.11	11.83	30.55	42.77	$\neg$
Federated DA	FADA [15]	21.00	$\emptyset$	62.98	272.90	$\neg$
	KD3A [16]	11.57	69.42	173.55	219.83	$\neg$
	FligHT	8.29	49.74	$\emptyset$	$\emptyset$	66.32

[11] C.-H. Yao, B. Gong, H. Qi, Y. Cui, Y. Zhu, and M.-H. Yang, "Federated multi-target domain adaptation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1424–1433.

[12] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[13] K. Zhang, M. Gong, and B. Schölkopf, "Multi-source domain adaptation: A causal view," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, 2015.

[14] W.-G. Chang, T. You, S. Seo, S. Kwak, and B. Han, "Domain-specific batch normalization for unsupervised domain adaptation," in *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, 2019, pp. 7354–7362.

[15] X. Peng, Z. Huang, Y. Zhu, and K. Saenko, "Federated adversarial domain adaptation," in *International Conference on Learning Representations*, 2020.

[16] H. Feng, Z. You, M. Chen, T. Zhang, M. Zhu, F. Wu, C. Wu, and W. Chen, "Kd3a: Unsupervised multi-source decentralized domain adaptation via knowledge distillation." in *ICML*, 2021, pp. 3274–3283.

[17] J. Liu, H. Xu, L. Wang, Y. Xu, C. Qian, J. Huang, and H. Huang, "Adaptive asynchronous federated learning in resource-constrained edge computing," *IEEE Transactions on Mobile Computing*, 2021.

[18] X. Ouyang, Z. Xie, J. Zhou, G. Xing, and J. Huang, "Clusterfl: A clustering-based federated learning system for human activity recognition," *ACM Transactions on Sensor Networks*, vol. 19, no. 1, pp. 1–32, 2022.

[19] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "Wifi csi based passive human activity recognition using attention based blstm," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2018.

[20] B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen, and M. Wu, "Two-stream convolution augmented transformer for human activity recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, 2021, pp. 286–293.

[21] W. Meng, Z. Liu, B. Li, W. Cui, J. T. Zhou, and L. Zhang, "Graphar: A lightweight human activity recognition model by exploring the subcarrier correlations," *IEEE Transactions on Wireless Communications*, 2023.

[22] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-free wireless localization and activity recognition: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6258–6267, 2016.

[23] B. Li, W. Cui, Y. Chen, J. T. Zhou, Z. Chen, Y. Li, and W. Min, "Cost-effective elderly fall detection with symmetry transformer networks," in *20th Int Conf on Smart City*. IEEE, 2022, pp. 1837–1845.

[24] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *Proceedings of the 17th annual international conference on mobile systems, applications, and services*, 2019, pp. 313–325.

[25] W. Cui, L. Zhang, B. Li, Z. Chen, M. Wu, X. Li, and J. Kang, "Semi-supervised deep adversarial forest for cross-environment localization," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 10 215–10 219, 2022.

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

[27] 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, pp. 7167–7176.

[28] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (gdpr)," *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, vol. 10, no. 3152676, pp. 10–5555, 2017.

[29] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.

[30] S. Itahara, T. Nishio, Y. Koda, M. Morikura, and K. Yamamoto, "Distillation-based semi-supervised federated learning for communication-efficient collaborative training with non-iid private data," *IEEE Transactions on Mobile Computing*, vol. 22, no. 1, pp. 191–205, 2021.

[31] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.

[32] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *International Conference on Machine Learning*. PMLR, 2020, pp. 5132–5143.

[33] Y. Tan, G. Long, L. Liu, T. Zhou, Q. Lu, J. Jiang, and C. Zhang, "Fedproto: Federated prototype learning across heterogeneous clients," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022, pp. 8432–8440.

[34] R. Presotto, G. Civitarese, and C. Bettini, "Fedclar: Federated clustering for personalized sensor-based human activity recognition," in *2022 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 2022, pp. 227–236.

[35] Y. Li, X. Wang, and L. An, "Hierarchical clustering-based personalized federated learning for robust and fair human activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 1, pp. 1–38, 2023.

[36] H. Yu, Z. Chen, X. Zhang, X. Chen, F. Zhuang, H. Xiong, and X. Cheng, "Fedhar: Semi-supervised online learning for personalized federated human activity recognition," *IEEE Transactions on Mobile Computing*, 2021.

[37] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. Springer, 2016, pp. 597–613.

[38] R. Li, Q. Jiao, W. Cao, H.-S. Wong, and S. Wu, "Model adaptation: Unsupervised domain adaptation without source data," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9641–9650.

[39] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation," in *International Conference on Machine Learning*. PMLR, 2020, pp. 6028–6039.

[40] J. Liang, D. Hu, Y. Wang, R. He, and J. Feng, "Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8602–8617, 2021.

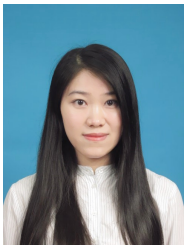
[41] N. Guha, A. Talwalkar, and V. Smith, "One-shot federated learning," *arXiv preprint arXiv:1902.11175*, 2019.

[42] M. Fredrikson, E. Lantz, S. Jha, S. Lin, D. Page, and T. Ristenpart, "Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing," in *23rd {USENIX} Security Symposium ({USENIX} Security 14)*, 2014, pp. 17–32.

[43] Y. Zhang, R. Jia, H. Pei, W. Wang, B. Li, and D. Song, "The secret revealer: Generative model-inversion attacks against deep neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 253–261.

[44] R. Müller, S. Kornblith, and G. E. Hinton, "When does label smoothing help?" *Advances in neural information processing systems*, vol. 32, 2019.
[45] A. Krause, P. Perona, and R. Gomes, "Discriminative clustering by regularized information maximization," *Advances in neural information processing systems*, vol. 23, 2010.
[46] S. Zhao, G. Wang, S. Zhang, Y. Gu, Y. Li, Z. Song, P. Xu, R. Hu, H. Chai, and K. Keutzer, "Multi-source distilling domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 975–12 983.
[47] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," *Advances in neural information processing systems*, vol. 32, 2019.
[48] W. B. Johnson and J. Lindenstrauss, "Extensions of lipschitz mappings into a hilbert space," *Contemporary Mathematics*, vol. 26, no. 189-206, p. 1, 1984.
[49] A. Baha'A, M. M. Almazari, R. Alazrai, and M. I. Daoud, "A dataset for wi-fi-based human activity recognition in line-of-sight and non-line-of-sight indoor environments," *Data in Brief*, vol. 33, p. 106534, 2020.
[50] D. Anguita, A. Ghio, L. Oneto, X. Parra, J. L. Reyes-Ortiz *et al.*, "A public domain dataset for human activity recognition using smartphones." in *Esann*, vol. 3, 2013, p. 3.

Bing Li is currently a Professor with the University of Electronic Science and Technology of China. He was a Senior Scientist at A*STAR Centre for Frontier AI Research (CFAR), Singapore, and Research Associate with the University of New South Wales, Australia. He received Ph.D. degree from Northeastern University, China, in 2018. His research interests include intelligent information extraction, natural language processing, and machine learning. He is the awardee of the 2019 ACM China SIGMOD Chapter Doctoral Dissertation Award. He served as area chair for ACL-ARR and PC members in various first-tier conferences, such as ICML, NeurIPS, AAAI, ICDE and ICLR.



Wei Cui received the M.E. and Ph.D. degrees in pattern recognition and intelligent system from Northeastern University, Shenyang, China, in 2013 and 2017, respectively. Currently, she is a senior scientist at Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. Her research interests include wireless sensor networks, wireless sensing, and machine learning. She serves as associate editor of Neurocomputing and Computers and Electrical Engineering. She also served as PC members in various first-tier conferences, such as CVPR and IJCAI. She is currently the Secretary of IEEE Sensors Council Singapore Chapter.

Le Zhang received the B.Eng. degree from the University of Electronic Science and Technology of China in 2011 and the M.Sc. and Ph.D. degrees from Nanyang Technological University (NTU), in 2012 and 2016, respectively. Currently, he is a Professor with the University of Electronic Science and Technology of China. With integrated learning as the center, he has carried out cross-directional research in multiple directions, such as deep learning, computer vision, wireless perception, and medical image processing. He has published more than 70 high-level papers in related fields (e.g., IEEE TPAMI/ICJV/CVPR/NeurIPS/AAAI, etc.). He has been selected as one of the top 2% scientists in the world published by Stanford University for many consecutive years. Serving as an invited editor for prestigious journals such as IEEE Trans on Neural Networks and Learning Systems, IEEE Trans on Big Data, Pattern Recognition, Neurocomputing, and so on.

Qi Yang received the M.E. degrees from Shandong University of Science and Technology, China, in 2024. His research interests include wireless sensor networks, wireless sensing, and federated learning.

Min Wu is currently a Principal Scientist at Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore. He received his Ph.D. degree in Computer Science from Nanyang Technological University (NTU), Singapore, in 2011 and B.E. degree in Computer Science from University of Science and Technology of China (USTC) in 2006. He received the best paper awards in IEEE ICIEA 2022, IEEE SmartCity 2022, InCoB 2016 and DASFAA 2015. He also won the CVPR UG2+ challenge in 2021 and the IJCAI competition on repeated buyers prediction in 2015. He has been serving as an Associate Editor for journals like Neurocomputing, Neural Networks and IEEE Transactions on Cognitive and Developmental Systems, as well as conference area chairs of leading AI and machine learning conferences, such as ICLR, NeurIPS, etc. His current research interests focus on AI and machine learning for time series data, such as deep learning, self-supervised learning, domain adaptation, and knowledge distillation for time series data.

Joey Tianyi Zhou is the Deputy Director and Principal Scientist, with the A*STAR Centre for Frontier AI Research (CFAR), Singapore. He is also holding a joint appointment with the Centre for Advanced Technologies in Online Safety (CATOS) as principal scientist. Before working at CFAR, he was a senior research engineer with SONY US Research Center in San Jose, USA. Dr.Zhou received a Ph.D. degree in computer science from Nanyang Technological University (NTU), Singapore. His current interests mainly focus on improving the efficiency and robustness of machine learning algorithms. In these areas, he has published more than 100 papers and received the Best Student Paper Nomination at the European Conference on Computer Vision (ECCV'16), Best Paper Award at IEEE SmartCity 2022, International Joint Conference on Artificial Intelligence (IJCAI) workshops, and Best Poster Award and runner-up prize at International Conference on Computer Vision (ICCV19) on HANDS workshop and its competition, respectively. Dr. Zhou regularly organizes workshops/tutorials at top-tier international conferences like CVPR, IJCAI, ICDCS, etc. He is serving on an Editorial Board for many leading journals like AIJ, IEEE Transactions, etc., and Area Chairs in top machine learning conferences like ICLR, ICML, NeurIPS, IJCAI, and Associate Programme Chair in IJCAI 2025, etc. He is listed in the Top 2% Scientists Worldwide 2023 by Stanford University.