Privacy-Preserving and Cross-Domain Human Sensing by Federated Domain Adaptation with Semantic Knowledge Correction

KAIJIE GONG, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, China
YI GAO, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, China
WEI DONG, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, China

Federated Learning (FL) enables distributed training of human sensing models in a privacy-preserving manner. While promising, federated global models suffer from cross-domain accuracy degradation when the labeled source domains statistically differ from the unlabeled target domain. To tackle this problem, recent methods perform pairwise computation on the source and target domains to minimize the domain discrepancy by adversarial strategy. However, these methods are limited by the fact that pairwise source-target adversarial alignment alone only achieves domain-level alignment, which entails the alignment of domain-invariant as well as environment-dependent features. The misalignment of environment-dependent features may cause negative impact on the performance of the federated global model. In this paper, we introduce FDAS, a Federated adversarial Domain Adaptation with Semantic Knowledge Correction method. FDAS achieves concurrent alignment at both domain and semantic levels to improve the semantic quality of the aligned features, thereby reducing the misalignment of environment-dependent features. Moreover, we design a cross-domain semantic similarity metric and further devise feature selection and feature refinement mechanisms to enhance the two-level alignment. In addition, we propose a similarity-aware model fine-tuning strategy to further improve the target model performance. We evaluate the performance of FDAS extensively on four public and a real-world human sensing datasets. Extensive experiments demonstrate the superior effectiveness of FDAS and its potential in the real-world ubiquitous computing scenarios.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**.

Additional Key Words and Phrases: Federated learning, Domain adaptation, Human sensing, Adversarial alignment.

ACM Reference Format:
Kaijie Gong, Yi Gao, and Wei Dong. 2024. Privacy-Preserving and Cross-Domain Human Sensing by Federated Domain Adaptation with Semantic Knowledge Correction . *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 1, Article 6 (March 2024), 26 pages. https://doi.org/10.1145/3643503

1 INTRODUCTION

In recent years, more and more smart distributed devices, such as smartphones, smart personal assistants, and other wearable devices have emerged. As the computing power of devices continues to increase and data privacy concerns rise, there is a growing trend toward keeping data and computation on these devices, especially in the area of human sensing [6, 16, 33, 39]. Federated learning (FL), such as FedAvg [27], provides a mechanism for local model training and global model updating to keep data locally on the devices. The federated paradigm improves data privacy and training efficiency in machine learning and has received widespread attention [6, 18, 32].

Yi Gao and Wei Dong are corresponding authors.
Authors' addresses: Kaijie Gong, The State Key Laboratory of Blockchain and Data Security, Zhejiang University , HangZhou, China , gongkj@zju.edu.cn; Yi Gao, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, HangZhou, China, gaoyi@zju.edu.cn; Wei Dong, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, HangZhou, China, dongw@zju.edu.cn.

2474-9567/2024/3-ART6 $15.00
https://doi.org/10.1145/3643503

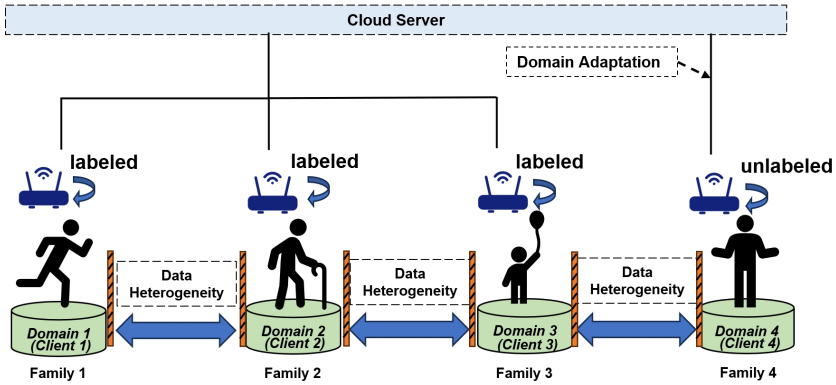


Fig. 1. The UFDA problem. We consider each domain as a separate client. The more detailed real-world scenarios (Scenario 1: each domain is treated as an independent client; Scenario 2: each client contains multiple domains; Scenario 3: each domain contains multiple clients.) are discussed in Section 8.2.4.

While promising, it is challenging to enable the federated global model trained on the decentralized clients to a new client with unlabeled data and achieve high performance. The performance degradation is attributed to domain shift [12, 21, 51] between the unlabeled client and the labeled clients. The problem is defined as Unsupervised Federated Domain Adaptation (UFDA) [37]. The UFDA and data heterogeneous under FL problem are complementary (see Figure 1). The former focuses on domain adaptation between multiple labeled domains and an unlabeled domain, while the latter aims to mitigate data heterogeneity among the labeled domains. However, if the target domain has a sufficient number of labels, the UFDA problem will degrade into data heterogeneity under FL.

To tackle the UFDA problem, recent methods FADA [37] and EI [17] utilize adversarial learning to minimize pairwise source-target domain discrepancy, thereby enhancing the capacity of the global model to incorporate more knowledge from the target domain. These methods are expected to map all the source and target features to a common and closer feature subspace by adversarial alignment. While adversarial alignment has achieved encouraging results in the UFDA setting, existing methods are limited by the fact that pairwise source-target adversarial alignment alone only achieves domain-level alignment, which entails the alignment of domain-invariant as well as environment-dependent features. The misalignment of environment-dependent features may cause a negative impact on the performance of the federated global model. For example, the aligned features may contain similar interference caused by tables around humans in the WIFI-based human sensing tasks [9, 44, 48]. Thus, the adversarial alignment of these environment-dependent features can minimize pairwise source-target domain discrepancy, but may result in negative gains for the federated global model.

In this paper, we propose a Federated adversarial Domain Adaptation with Semantic Knowledge Correction method (FDAS), which aims to correct source and target features to have more semantic knowledge during the pairwise source-target adversarial alignment. To be specific, we achieve domain-level and semantic-level alignment simultaneously to improve the semantic quality of the aligned features thereby reducing the misalignment of environment-dependent features. However, there exist significant barriers to achieving effective domain-level and semantic-level alignment in the UFDA setting.

(1): Irrelevant source domains may cause the negative transfer problem [35]. To mitigate the negative transfer, existing works have explored the client selection [33] and dynamic aggregation weights [10, 37] to

reduce the influence of irrelevant source domains. However, client selection or dynamic aggregation weights methods are based on domain granularity. Thus, these methods will affect the entire domain features, limiting the contribution of features that are more readily adaptable to the target domain. We propose a federated feature selection mechanism that differs from existing domain granularity methods and operates at a finer feature granularity. It selects easy-to-transfer source features first and then utilizes the learned knowledge to facilitate the transfer of other hard-to-transfer features. The selected easy-to-transfer features possess greater similarity with the target domain and have more contributions to the target domain, thereby reducing negative transfer.

(2): It is difficult to address the misalignment of samples belonging to different class labels but with high similarity or the samples belonging to the same class label but with low similarity . One possible solution is to learn an embedding subspace for each source domain, where features with the same class labels are clustered together and features with different class labels are separated [31]. However, this method may cause features with the same class labels to be pulled far away from each other across different domains, which is not conducive to transfer. We propose federated feature refinement mechanism, which leverages global semantic knowledge to reinforce the semantic components of the features.

(3): The bias accumulation caused by the incorrect conditional information . Conditional Adversarial alignment [19, 26] can alleviate the problem of mode collapse [30] by the guided conditional information. However, these approaches are based on a strong pre-assumption that the bias caused by incorrect conditional information can be reduced by subsequent correct conditional information. Furthermore, the bias can easily accumulate and be magnified in the UFDA setting due to the possibility of mutual influence among this incorrect conditional information during different pairwise source-target computations. The easy-to-transfer features selected by our federated feature selection mechanism often possess higher confidence and are more likely to be correctly classified due to their ease of transferability. Thus, our selected features can enable the generation of higher-quality conditional information.

(4): The generation of high-quality semantic knowledge is challenging during the pairwise source-target semantic-level alignment. The semantic knowledge refers to prototypes [20, 29, 47], which can capture similar semantic knowledge from different samples. However, prototypes are limited by the specific domains and are difficult being applied to the target domain. We design cross-domain contrastive correction loss to generate domain-invariant prototypes.

Thus, we develop Semantic-guidance Contrastive Correction Strategy (SGCC) and Selection-refinement Adversarial Correction Strategy (SRAC) to achieve enhanced semantic-level and domain-level alignment simultaneously. We implement a prototype of FDAS on multiple heterogeneous devices and conduct extensive experiments on four public human sensing datasets and a real-world human sensing dataset, including WIFI, images, IMU, UWB and smartphone. The major contributions of this work can be summarized as follows:

- We propose a Federated adversarial Domain Adaptation with Semantic Knowledge Correction method (FDAS) to address the UFDA problem. We develop Semantic-guidance Contrastive Correction Strategy (SGCC) and Selection-refinement Adversarial Correction Strategy (SRAC) to correct pairwise source-target computation to achieve domain-level and semantic-level alignment simultaneously.
- We design a cross-domain semantic similarity metric as well as feature selection and refinement mechanisms for SGCC and SRAC to enhance the domain-level and semantic-level alignment. The feature selection mechanism can alleviate the negative transfer problem and reduce the bias accumulation caused by incorrect conditional information. The feature refinement mechanism can address the misalignment of samples belonging to different classes but with high similarity and protect feature privacy.
- We conduct comprehensive experiments on four public human sensing datasets and a real-world human sensing dataset. Extensive experiments demonstrate the superior effectiveness of FDAS and its potential in the real-world ubiquitous computing scenarios.

2 RELATED WORK

2.1 Cross-domain Human Sensing

Since adopting machine learning approaches as a large-scale solution for human sensing recognition tasks still face a major challenge that the performance of a machine-learned model may drop significantly when being deployed in a new environment, many researchers have tried to address the cross-domain issue. Widar3.0 [53] leverages multiple WIFI links to extract domain-independent features at the lower signal level [50] [43] [24], which is called BVP. However, Widar3.0 can only be applied to WIFI scenarios with at least 3 links. Gao et al. [12] extract position-independent features, denoted as MNP, from the hand-oriented view. However, it only can be suitable to cross-position model deployment. EI [17] employ domain adversarial network to align the features between source domains and target domains in the feature spaces. CrossSense [51] presents a translation solution to generate virtual samples for new environment and train a robust model, WiGr [52] uses the similarity of the target domain sample features to the source domain sample features to reduce the impact of domain-specific features. However, most of these methods require access to raw data, which is not allowed in the federated setting. To the best of our knowledge no prior works have investigated cross-domain human sensing in the federated setting.

2.2 Unsupervised Federated Domain Adaptation

Unsupervised Federated Domain Adaptation (UFDA) Methods achieve multi-source domain adaptation while maintaining data privacy and training efficiency. The UFDA problem was proposed by Peng et al. [37], and there are mainly two works discussing the problem recently. FADA [37] is a federated domain adaptation method using domain adversarial and feature disentangling, FADA utilizes adversarial alignment to minimize the discrepancy between the target domain and the source domain and enhances knowledge transfer through feature disentanglement. It also designs dynamic weights to address the negative transfer problem. KD3A [10] is a knowledge distillation method. KD3A deviates from the federated learning paradigm and aims to achieve high-performance target model rather than a viable global model. KD3A does not incorporate federated training, thereby sacrificing the federated training efficiency. KD3A proposes a multi-source model knowledge distillation method based on Knowledge Vote to obtain high-quality domain consensus knowledge. Then, the obtained consensus knowledge is utilized to dynamically weigh the different source domains to prevent negative transfer. Finally, BatchNorm MMD is utilized to minimize the domain distribution distance.

Our method (FDAS) addresses the significant barriers when employing the adversarial strategy in the federated setting, which differs from KD3A completely. Compared with FADA, our method also has significant differences: (1) FADA only used the original adversarial alignment strategy [11] on pairs of source and target domains and ignored the challenges of adversarial alignment strategy in the federated setting. Our method designs feature selection and refinement mechanisms to optimize adversarial alignment strategy in the federated setting. (2) FADA learned domain-invariant features for each source domain by feature disentanglement. However, feature disentanglement is limited by additional time-consuming adversarial training and ignored the experiment-specific but task-related features. Our Method generates global semantic knowledge across domains and leverages all the shared knowledge to correct local features simultaneously have semantic-level and domain-level information to improve the semantic of the aligned features.

2.3 Adversarial Alignment

Adversarial alignment aims to generate transferable features by minimizing the discrepancy between source and target domains. DANN [11]first applied an adversarial network in the Unsupervised Domain Adaptation (UDA) problem to align the features between the source and target domains. CDAN [26] employed a conditional adversarial network to the UDA problem to alleviate the effects of mode collapse [28].

Our Selection-refinement Adversarial Correction Strategy approach (SRAC) differs from these methods in that: (1) Our method refines the source-target features with local semantic knowledge to address the misalignment of samples belonging to different classes but with high similarity and protect feature privacy. (2) Our method selects confident-to-transfer samples to generate higher-quality conditional information to alleviate negative transfer and reduce the bias accumulation caused by incorrect conditional information effectively. (3) Our method imposes diversity constraints on unconfident-to-transfer samples, so that unconfident-to-transfer samples still have the opportunity to be selected.

2.4 Prototype Contrastive Learning

Prototypes are usually defined as the mean feature of the same class [29, 40]. Since prototypes have the ability to capture similar semantic knowledge from different samples, prototype learning is widely used in transfer learning [38]. Contrastive Learning [5, 49]aims to achieve embeddings where the same samples are pulled closer and those of different samples are pushed apart. FedPCL [41] utilized prototype contrastive learning to extract more shared features in the federated setting.

Our Semantic-guidance Contrastive Correction Strategy (SGCC) is different from FedPCL: (1) FedPCL aggregated all local features on average as a global federated prototype. We consider that different local features have different contributions to the global federated prototype depending on how semantic similarity between local features and the global federated prototype. Thus, we design a semantic similarity-aware local features selection method to enable the global prototype contains more cross-domain semantic information. (2) The contrastive alignment in FedPCL is based on the guidance with true labels. Our method designs contrastive alignment with unlabeled data.

3 MOTIVATION

In this section, we provide a detailed definition of the Unsupervised Federated Domain Adaptation (UFDA) problem and our key insights to address the problem.

3.1 Motivating Use Case

We consider a typical ubiquitous computing scenario: WiFi-based gesture recognition (see Figure 1). A significant impediment in training gesture recognition models lies in adapting the model to new unseen users. Collecting labeled data from unseen users is extremely challenging, primarily due to privacy considerations and the expensive labeling. Furthermore, the collected seen user data may be owned by different families or confidential entities, carrying a substantial amount of user privacy. Therefore, the goal of the UFDA problem is whether it is possible to attain a high-performance cross-user gesture recognition model while safeguarding the privacy of both seen and unseen users.

3.2 Problem Definition

Lets $\mathcal{D}_S$ and $\mathcal{D}_T$ denote the source domain and target domain. The data owned by the source and target domains have the same label distribution but different feature distributions. In the UFDA problem, we have K source domains $\{\mathcal{D}_S^k\}_{k=1}^K$ and one target domain $\mathcal{D}_T$. For the $k \in K$ source domain, we have N_k labeled samples $\{(X_i^k, Y_i^k)\}_{i=1}^{N_k}$. For the target domain, we have N_T unlabeled samples $\{X_i^T\}_{i=1}^{N_T}$. Following the federated paradigm, we have a global feature extractor $\mathcal{F}$ and a global classifier C. For the $k \in K$ source domain, we have a domain-specific feature extractor $\mathcal{F}_S^k$, a domain-specific classifier C_S^k and a domain-specific discrimintor $\mathcal{DI}_S^k$. For the target domain, we have a feature extractor $\mathcal{F}_T$ and a classifier C_T. The goal of UFDA is to leverage knowledge learned from muti-source domains$\{\mathcal{D}_S^k\}_{k=1}^K$ and learn an optimal feature extractor $\mathcal{F}_T$ and classifier C_T for $\mathcal{D}_T$.

Method	user14 -> user13	user15 -> user13	user16 -> user13	user14,15,16 -> user13
source-only	73.33	74	75.33	78
alignment	80.67	79.33	75.33	76

Table 1. The target accuracy of domain-level alignment and global domain optimization. Source-only method: the combination of source domains and trains a single model without feature alignment; Alignment method: domain-level adversarial alignment. The "->" symbol represents the direction of transfer learning, where the source domains are listed before the arrow and the target domain is listed after the arrow.

3.3 Pairwise Domain-level Alignment and Global Model Optimization

Existing work [37] has proved the effectiveness of pairwise domain-level alignment in the UFDA setting. However, single domain-level alignment may cause a negative impact on the federated global model optimization. Pairwise domain-level alignment refers to the process of training a source model for pairwise source-target domains to reduce the distributional differences between the source and target domains. Global model optimization involves the federated aggregation of the source models to further minimize the distributional discrepancies between the source domains and the target domain.

To investigate a fine-grained analysis of the gap between domain-level alignment and global model optimization, we conduct experiments on the Widar3.0 (BVP) datasets [53]. Our feature extractor consists of two convolutional layers and three fully connected layers. We use a fully connected layer as the classifier and two fully connected layers as the discriminator. The learning rates for our feature extractor, classifier, and discriminator have been configured at $1e^{-4}$. We set the local epoch is 1 and the global epoch is 400. We selected three users as the source domains and one user as the target domain. The results are presented in Table 1. The source-only method is to directly deploy the model trained in the source domains to the target domain. The domain-level alignment method is to reduce the discrepancy between the source and target domains through an adversarial alignment strategy. Global model optimization involves averaging all locally optimized models using FedAvg [27].

As is shown in Table 1, we can improve the accuracy of a single domain-level alignment by utilizing adversarial strategies, such as pairwise domain-level alignment from a source domain (e.g. user14) to a target domain (e.g. user13). Nevertheless, the accuracy of the federated global model decreases at this point. This is because domain-level alignment can only guarantee the similarity of pairwise source-target features, but this similarity maybe only environment-dependent and cannot provide effective gains for the global model. Thus, the observation guided us to conclude that the domain-level alignment is ambiguous and does not entail the alignment of features relevant to human activity recognition.

One direct solution to address the misalignment of environment-dependent features is to extract domain-invariant features for each source domain to enhance transferability. However, despite considerable exploration [8, 34, 36, 37], this approaches remain highly challenging and result in the loss of certain environment-specific features that may be beneficial for human activity recognition, such as the co-occurrence of "push & pull" with "door". To track the issue, we have proposed an approach that does not rely on the direct extraction of domain-invariant features. Instead, we propose a method for simultaneous alignment at both the domain level and semantic level. The semantic-level alignment exploits the shared semantic information (task-related) within each domain, while domain-level alignment facilitates the extraction of domain-invariant features. The semantic and domain-invariant features are highly likely to pertain to the experiment-specific but task-related features.

3.4 The Negative Transfer Problem

Irrelevant source domains, which are very different from the target domain, may cause the negative transfer problem [35]. We set up four different federated domains to analyze the phenomenon of negative transfer (see

Method	user 14,15 -> user 13	user 14,16-> user 13	user 15,16 -> user 13	user 14, 15,16 -> user 13
source-only	72.67	77.33	70.00	79.79
alignment	78.66	81.33	68.66	83.33

Table 2. The target accuracy of different UFDA setting. Source-only method: the combination of source domains and trains a single model without feature alignment; Alignment method: domain-level adversarial alignment. The "->" symbol represents the direction of transfer learning, where the source domains are listed before the arrow and the target domain is listed after the arrow.

Table 2). The experiment setup is identical to that described in Section 3.3. Table 2 demonstrates that there are substantial variations in the target accuracy achieved by different combinations of source domains, particularly in the transfer from `user15` and `user16` to `user13`, where the model adaptation method results in a negative gain. Therefore, the presence of irrelevant source domains that are very different from the target domain or even some malicious source domains may lead to negative transfer, resulting in decreased target model performance.

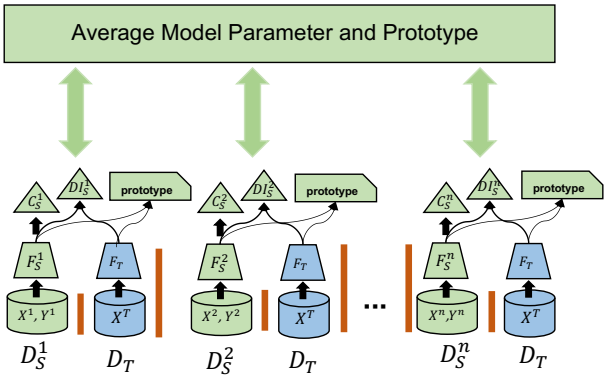
To mitigate the negative transfer, an intuitive idea is to adopt client selection methods, such as similarity-based node selection [33]. However, directly excluding irrelevant source domains through client selection methods can minimize negative transfer, but it also results in a significant reduction of source data, which leads to a decrease in target model performance. As is shown in Table 2, `user15` contributes the least gain compared to `user14` and `user16`. However, directly discarding `user15` would lead to performance degradation. Another solution is to adopt dynamic aggregation weights based on clustering-based contribution measuring [37] and knowledge voting [10], to reduce the influence of irrelevant source domains. However, dynamic aggregation weights methods are based on domain granularity. Thus, the weights will affect the entire domain features, which limits the contribution of easy-to-transfer features within that domain. FADA [37] and KD3A [10] employ dynamic aggregation weight methods. we will provide relevant comparative results in Section 7.

In this paper, we propose a federated feature selection mechanism that operates at a finer granularity compared to client selection and dynamic aggregation weights methods. By selecting easy-to-transfer features, which have more contributions to the target domain, our approach can mitigate negative transfer while simultaneously enhancing the contributions of irrelevant source domains. Our feature selection mechanism incorporates the idea of curriculum learning, which prioritizes the easier-to-transfer features and gradually incorporates the learned knowledge to transfer more challenging features.

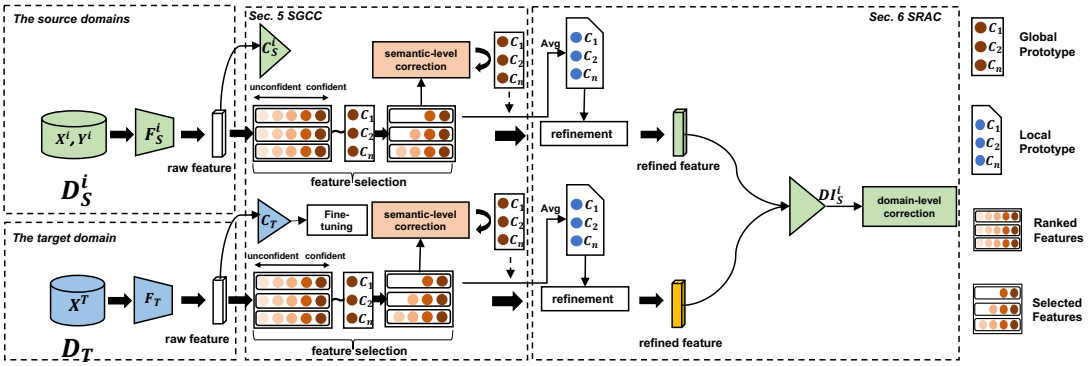
4 APPROACH OVERVIEW

To solve the UFDA problem, we propose a Federated adversarial Domain Adaptation with Semantic knowledge correction method (FDAS). The overall framework of FDAS is depicted in Fig. 2(a), which requires pairwise source-target domain computation. Specifically, each source domain consists of a domain-specific feature extractor, domain-specific classifier, and domain discriminator. The domain discriminator is utilized to minimize the discrepancy between source and target domains.

In each federated training round, each source domain $\mathcal{D}_S^k$ first downloads a updated global feature extractor $\mathcal{F}$ and classifier C to train a domain-specific feature extractor $\mathcal{F}_S^k$, a domain-specific classifier C_S^k and send $\mathcal{F}_S^k$, C_S^k to the target domain for pairwise computation. To achieve domain-level alignment, the target domain needs to transmit the refined features to the source domain during the pairwise computation. Note that our feature refinement leverages local target prototypes to obfuscate raw target features, thereby avoiding the direct transmission of raw target features. Then, each pairwise computation between the source$\mathcal{D}_S^k$ and target domain $\mathcal{D}_T$ perform Semantic-guidance Contrastive Correction Strategy (SGCC) and Selection-refinement Adversarial



(a) FDAS



(b) SGCC and SRAC

Fig. 2. (a) The framework of the proposed FDAS. It requires pairwise source-target domain computation and achieves domain-level and semantic-level alignment simultaneously. (b) One of the pairwise source-target domain computations contains SGCC and SRAC. SGCC consists of feature selection and semantic-level correction components. SRAC consists of feature selection, feature refinement and domain-level correction components. We authorize the exchange of model parameters and prototypes between clients and servers. We only allow the transmission of target refined features between pairwise source-target computation. The feature refinement mechanism can protect the target feature privacy.

Correction Strategy (SRAC) to achieve domain-level and semantic-level alignment simultaneously, which is depicted in Fig. 2(b).

SGCC first performs feature selection to select reliable features to generate a global prototype $\mathbb{P} = \{\overline{\mathcal{P}}^j\}_{j=1}^{N_p}$, where N_p is the number of categories in this task. SGCC then performs semantic-level correction based on the global prototype to correct local features to carry more semantic-level information. Meanwhile, SRAC first uses the results of feature selection to exclude unconfident-to-transfer features and perform feature refinement with the global prototype learned from SGCC. SRAC then combines the refined features of the source-target domains and the output of the classifier as the input to the discriminator to perform the domain-level correction to correct local features to carry more domain-level information.

SGCC and SRAC update iteratively and jointly. SGCC provides SRAC with global semantic knowledge guidance to help SRAC achieve better domain-level alignment. SRAC provides cross-domain knowledge support for SGCC and enable it to achieve better semantic-level alignment.

After each round, each source domain uploads the local optimized model and aggregates all models on the server. Until the end of the federation training, the target domain downloads the optimized global model from the server. Under our optimization, the global prototype will be domain-invariant. Thus, we use high-quality pseudo-labels generated by the semantic similarity measurements between the local features and global prototype to fine-tune the target domain model and further improve the target model performance.

5 SEMANTIC-GUIDANCE CONTRASTIVE CORRECTION STRATEGY

The object of Semantic-guidance Contrastive Correction Strategy (SGCC) is to provide global semantic knowledge and perform semantic-level correction.

Global semantic knowledge generative. SGCC uses prototypes to represent semantic knowledge because prototypes can capture semantic similarity from different samples across domains [42, 46]. However, The prototypes are usually aggregated directly from all features, which may have the following challenges in the UFDA problem: (1) Due to feature shift, different local source features have different contributions to the global prototype. Thus, it is not wise to directly aggregate all features on average. (2) Prototypes are limited by the specific domains and are difficult to apply to the target domain. Thus, generating high-quality prototypes for the target domain is challenging.

To address the challenges, SGCC uses feature selection mechanisms to select reliable features before aggregation. Inspired by PFAN [4], we desgin a feature selection based on cross-domain semantic similarity metric for each source domain to achieve high-quality domain-invariant prototypes. We define the cross-domain semantic similarity Φ for each source domain $\mathcal{D}_S^k$, where Φ is computed as follows:

$$\Phi_{(X_i^{k(j)})} = \mathrm{Sim}(\mathcal{F}_S^k(X_i^{k(j)}), \overline{\mathcal{P}}_{old}^j) + \mathrm{Sim}(\mathcal{F}_S^k(X_i^{k(j)}), \overline{\mathcal{P}}_{T(old)}^j), \tag{1}$$

where $\mathrm{Sim}(.,.)$ denotes the cosine similarity, $\overline{\mathcal{P}}_{old}^j$ denotes the global prototype belonging to class j in the previous round, $\overline{\mathcal{P}}_{T(old)}^j$ denotes the local target prototype belonging to class j in the previous round, $X_i^{k(j)}$ represents the i-th sample from the k-th source domain belonging to class j. In the early stage of training, the first term plays a dominant role because the large discrepancy between the source and target domain. In the later stage of training, the second term gradually plays a greater role because the certain source-target discrepancy reduction. For the target domain $\mathcal{D}_T$, Φ can be obtained by:

$$\Phi_{(X_i^{T(j)})} = \mathrm{Sim}(\mathcal{F}_T(X_i^{T(j)}), \overline{\mathcal{P}}_{T(old)}^j), \tag{2}$$

where $X_i^{T(j)}$ represents the i-th sample from the target domain belonging to class j. We select the top-N reliable sample features of each source domain according to the value of Φ and compute the local source prototype $\overline{\mathcal{P}}_k^j$ on the selected features$\hat{\mathcal{D}}_S^{k(j)}$:

$$\overline{\mathcal{P}}_k^j = \frac{1}{|\hat{\mathcal{D}}_S^{k(j)}|} \sum_{X_i^{k(j)} \in \hat{\mathcal{D}}_S^{k(j)}} \mathcal{F}_S^k(X_i^{k(j)}). \tag{3}$$

For the target domain, we set a more stringent condition to ensure that the selected sample features of the target domain can provide effective cross-domain semantic contributions. We constrain that the features can be selected only when the value of Φ exceeds a certain threshold τ. So, we can obtain the local target prototype $\overline{\mathcal{P}}_T^j$

on the selected features $\hat{\mathcal{D}}_T^{(j)}$ as follows:

$$\overline{\mathcal{P}}_T^j = \frac{1}{|\hat{\mathcal{D}}_T^{(j)}|} \sum_{X_i^{T(j)} \in \hat{\mathcal{D}}_T^{(j)}} \mathcal{F}_T(X_i^{T(j)}). \tag{4}$$

Thus, our feature selection always tends to select features with higher global semantic similarity. After feature selection, SGCC averagely aggregates the local prototypes of all the source and target domains on the server to generate global semantic knowledge:

$$\overline{\mathcal{P}}^j = \frac{\sum_{k \in [0,N_k)} (|\hat{\mathcal{D}}_S^{k(j)}|.\overline{\mathcal{P}}_k^j + |\hat{\mathcal{D}}_T^{(j)}|.\overline{\mathcal{P}}_T^j)}{\sum_{k \in [0,N_k)} (|\hat{\mathcal{D}}_S^{k(j)}| + |\hat{\mathcal{D}}_T^{(j)}|)}, \tag{5}$$

where N_k is the number of source domains. In the early training stage, the contribution of the target local prototype $\overline{\mathcal{P}}_T^j$ is small since only a few target domain features are selected. However, our SRAC can achieve domain-level alignment and more target domain features will be selected progressively. In this way, our global prototype will capture more semantic features that are shared across all source and target domains.

Semantic-level correction. To leverage global semantic knowledge to correct local features to carry more semantic knowledge, we define the cross-domain prototype contrastive correction loss term for the source domain as follows:

$$\mathcal{L}_{scon} = -\alpha \log \frac{\exp(\mathrm{Sim}(Z_i^{k(j)}, \overline{\mathcal{P}}^j)/\tau)}{\sum_{j_a \in C(j)} \exp(\mathrm{Sim}(Z_i^{k(j)}, \overline{\mathcal{P}}^{j_a})/\tau)} - (1-\alpha) \log \frac{\exp(\mathrm{Sim}(Z_i^{k(j)}, \overline{\mathcal{P}}_T^j)/\tau)}{\sum_{j_a \in C(j)} \exp(\mathrm{Sim}(Z_i^{k(j)}, \overline{\mathcal{P}}_T^{j_a})/\tau)}, \tag{6}$$

where $Z_i^{k(j)} = \mathcal{F}_S^k(X_i^{k(j)})$, $C(j) = \{j_a \in [1, C], j_a \neq j\}$ is the set of labels distinct from j, τ is the temperature that can adjust the tolerance for feature difference, α is the weighting parameters. The first term aims to align the source features to the global prototype. The second term aims to align the source feature to the local target prototype. For the target domain, we define confident-prototype contrastive correction loss term as follows:

$$\mathcal{L}_{tcon} = -\log \frac{\exp(\mathrm{Sim}(Z_i^{T(j)}, \overline{\mathcal{P}}^j)/\tau)}{\sum_{j_a \in C(j)} \exp(\mathrm{Sim}(Z_i^{k(j)}, \overline{\mathcal{P}}^{j_a})/\tau)}, \tag{7}$$

where $Z_i^{T(j)} = \mathcal{F}_T(X_i^{T(j)})$ and $X_i^{T(j)} \in \hat{\mathcal{D}}_T^{(j)}$.

The motivation for our modification to the classic prototype contrastive alignment [41] in our UFDA setting is to enhance higher-quality domain-invariant prototypes across the source and target domains: (1) For the source domain, we align local features to the target domain in the hope that the subsequent global prototype will contain more cross-domain features and have a greater chance of being selected by SRAC as confident-to-transfer features, which is depicted in section 5. (2) For the target domain, we only align the features selected by feature selection to global prototype in order to reduce the impact of misclassified target features.

6 SELECTION-REFINEMENT ADVERSARIAL CORRECTION STRATEGY

In this section, we introduce Selection-refinement Adversarial Correction Strategy(SRAC). The object of SRAC is to utilize global semantic knowledge learned from SGCC, and perform domain-level correction. The domain-level correction is based on an adversarial strategy. We have discussed the barriers to applying the adversarial strategy to the UFDA setting in section 1. Thus, we devise federated feature selection and refinement mechanisms to optimize adversarial strategy in the UFDA setting.

Feature selection. Our feature selection mechanism aims to address the following two challenges: (1) Traditional adversarial alignment, typically aims to align pairwise source-target features, to minimize the distribution discrepancy between the two domains. It is highly challenging because we cannot guarantee that the features of the two domains can be aligned correctly in the feature space. The misalignment of irrelevant features may result in negative transfer especially when considering the presence of malicious domains. (2) Conditional adversarial alignment can alleviate the problem of mode collapse by the guided conditional information. The conditional information is generally defined as the output vector of the classifier. However, the output vector may be the result of misclassification and give wrong guidance for adversarial alignment.

Our intuition is that source-target features with high similarity often possess higher transferability and are more likely to be classified correctly. Thus, Higher transferability implies less negative transfer, while higher classification accuracy indicates higher-quality conditional information.

Therefore, we perform feature selection before domain-level alignment. Specifically, SRAC shares the feature section mechanism with SGCC and only selects features with high similarity with the global prototype to participate in adversarial alignment each round. The selected features will lead to a more confident output of the classifier as conditional information, thereby reducing the bias caused by incorrect conditional information. We name the selected features as confident-to-transfer features and the other unselected features as unconfident-to-transfer features.

For the target domain, the confident-to-transfer features selected by SRAC are the same as those of SGCC. For the source domain, we compute $\Phi_{(X_i^{k(j)})} = \mathrm{Sim}(\mathcal{F}_S^k(X_i^{k(j)}), \overline{\mathcal{P}}_{old}^j)$ and also consider a more stringent threshold τ to limit the negative transfer of features below the threshold. We consider a slowly rising threshold to improve the quality of confident-to-transfer samples for all the source and target domains:

$$\tau = \frac{1}{1 + e^{-(round+1)/T}} - 0.05, \tag{8}$$

where *round* denotes the current training round and T is a constant, which is used to control the rate of change. Finally, we can obtain the target confident-to-transfer features $\hat{\mathcal{D}}_T^{(tr)}$ and the source confident-to-transfer features $\hat{\mathcal{D}}_S^{(tr)}$.

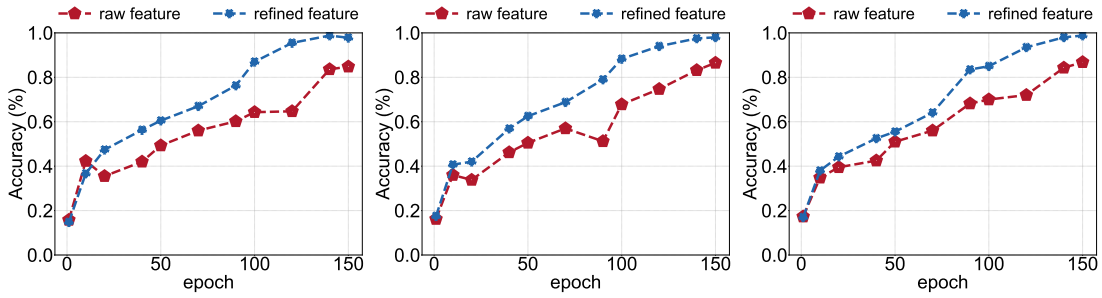
The main idea behind feature selection is inspired by the curriculum learning approach, where the model prioritizes learning easier knowledge first and then uses the acquired knowledge to learn more complex and challenging knowledge, which will be evaluated in section 7. In comparison with FADA, our feature selection mechanism can accelerate the convergence speed of adversarial alignment and reduce communication overhead. A detailed discussion of this is provided in section 7.

Feature refinement. Feature selection can exclude the negative effects of the unconfident-to-transfer features. However, even if we have correct conditional information, it is still difficult to reduce the misalignment of samples belonging to different classes but with high similarity and enhance the alignment of samples belonging to the same class but with low similarity.

Inspired by the effect of conditional information [7, 15, 23], we consider the additional semantic knowledge as a new alignment constraint. Specifically, we achieve local refined features by element-wise addition of the local features and the local prototypes. The main motivation is that we can enhance the similarity between the same class and reduce the similarity between different classes through refined features. In this way, feature refinement can enhance the adversarial alignment. In addition, the refined feature avoids the transmission of the raw features and protects feature privacy.

We can obtain the refined features $X_i^{k(j)'}$ for the source domain:

$$X_i^{k(j)'} = X_i^{k(j)} + \overline{\mathcal{P}}_k^j, \tag{9}$$



(a) user14 (b) user15 (c) user16

Fig. 3. The model accuracy on the raw and refined features. Raw features: the output of local feature extractor; refined features: element-wise addition of the local raw features and the local prototypes. datasets: widar3.0 (bvp). Source domain: (a)user14 (b)user15 (c)user16.

and the refined features $X_i^{T(j)'}$ for the target domain:

$$X_i^{T(j)'} = X_i^{T(j)} + \overline{\mathcal{P}}_T^j, \tag{10}$$

where $X_i^{k(j)} \in \hat{\mathcal{D}}_S^{k(tr)}$ and $X_i^{T(j)} \in \hat{\mathcal{D}}_T^{(tr)}$.

The core idea of our feature refinement is to leverage global semantic knowledge to reinforce the semantic components of the features. We evaluated the performance of global semantic knowledge and refined features to validate the potential of refined features (see Fig. 3). We observed that the accuracy of our refined features is consistently higher than that of the raw features in the later stage of training, indicating that the refined features contain more semantic knowledge compared to the raw features. Thus, in this way, some samples belong to different classes but have high similarity will be differentiated, while some samples that belong to the same class but have low similarity will be brought closer together.

Domain-level correction. To achieve domain-level alignment, we define the adversarial loss term based on feature selection and feature refinement mechanisms. Specifically, we can obtain conditional adversarial alignment on selected source features$\hat{\mathcal{D}}_S^{k(tr)}$ and selected target features $\hat{\mathcal{D}}_T^{(tr)}$through two steps for each source-target pair. In the first step, we keep the parameters of the feature extractor on the source-target pair unchanged and train the domain discriminator:

$$\mathcal{L}_{advDI} = -E_{X_i^{k'} \sim \hat{\mathcal{D}}_S^{k(tr)}}[\log \mathcal{DI}_S^k(Z_i^{k'} \oplus y_s^i)] - E_{X_i^{T'} \sim \hat{\mathcal{D}}_T^{k(tr)}}[\log(1 - \mathcal{DI}_S^k(Z_i^{T'} \oplus y_t^i))], \tag{11}$$

where $\oplus$ is the concatenation operation, $X_i^{k'} = \{X_i^{k(j)'}\}_{j=1}^C$ (C denotes the number of class labels), $X_i^{T'} = \{X_i^{T(j)'}\}_{j=1}^C$, $Z_i^{k'} = \mathcal{F}_S^k(X_i^{k'})$, $Z_i^{T'} = \mathcal{F}_T(X_i^{T'})$, $y_s^i = C_s^k \mathcal{F}_S^k(X_i^k)$ and $X_i^k \in \hat{\mathcal{D}}_S^{k(tr)}$, $y_t^i = C_s^k \mathcal{F}_T(X_i^T)$ and $X_i^T \in \hat{\mathcal{D}}_T^{(tr)}$. In the second step, we keep the parameters of the domain discriminator unchanged and train the local feature extractor:

$$\begin{aligned} \mathcal{L}_{advF_s} &= -E_{X_i^{k'} \sim \hat{\mathcal{D}}_S^{k(tr)}}[\log \mathcal{DI}_S^k(\mathcal{F}_S^k X_i^{k'}]; \\ \mathcal{L}_{advF_t} &= -E_{X_i^{T'} \sim \hat{\mathcal{D}}_T^{k(tr)}}[\log \mathcal{DI}_S^k(\mathcal{F}_T X_i^{T'})]. \end{aligned} \tag{12}$$

Next, we further enhance the domain-level correction by imposing the entropy minimization constraints [13] and diversity constraint [45] on the unselected features. The combination of these two constraints can increase the possibility of unconfident-to-transfer being selected in the later training stage.

7 TRAINING OPTIMIZATION

SGCC and SRAC update iteratively and jointly. SGCC provides SRAC with global semantic knowledge guidance to help SRAC achieve better domain-level alignment. SRAC provides cross-domain knowledge support for SGCC, enables semantic knowledge that is domain-invariant and achieves better semantic-level alignment. Thus, in each round, we define the total source domain training loss as follows:

$$\mathcal{L}_{s-total} = \alpha_s \mathcal{L}_{cls} + \beta_s \mathcal{L}_{scon} + (1 - \alpha_s - \beta_s)\mathcal{L}_{advF_s}, \tag{13}$$

where $\mathcal{L}_{cls}$ is the task-specific cross-entropy loss. We define the total target domain training loss as follows:

$$\mathcal{L}_{t-total} = \alpha_t \mathcal{L}_{tcon} + (1 - \alpha_t)\mathcal{L}_{advF_t} + \beta_t \mathcal{L}_{min} - (1 - \beta_t)\mathcal{L}_{div}, \tag{14}$$

where $\mathcal{L}_{min}$ is the entropy minimization constraints [13], $\mathcal{L}_{div}$ is the diversity constraints [45].

At the end of federated training, our gradually optimized global prototype will be domain-invariant. Thus, we further propose a similarity-aware model fine-tuning strategy based on the global prototype. Specifically, we first use the global classifier to generate a series of pseudo-labels. Then we share the feature selection mechanism with SGCC and SRAC to select features with high cross-domain semantic similarity ($\Phi > \tau$) and exclude low-quality pseudo-labels. Finally, we utilize high-quality pseudo-labels to train a personalized target domain model. The detailed training procedure is presented in Algorithm 1.

Algorithm 1 The Algorithm of FDAS

Input: K source domain $\{\mathcal{D}_S^k\}_{k=1}^K$, one target domain $\mathcal{D}_T$, K source feature extractor $\mathcal{F}_S^k$, K source classifier C_S^k, K source discriminator $\mathcal{DI}_S^k$, local source prototype $\overline{\mathcal{P}}_S^k$, local target prototype $\overline{\mathcal{P}}_T$, global feature extractor $\mathcal{F}$, global classifier C, global prototype $\overline{\mathcal{P}}$.

Output: Target domain feature extractor $\mathcal{F}_T$, classifier C_T

CenterExecutes:
initialize $\mathcal{F}$, C, $\overline{\mathcal{P}}$
for each round t = 1, 2, … **do**
  for each client $c \in K$ **do**
    $\mathcal{F}_S^k, C_S^k, \overline{\mathcal{P}}_S^k, \overline{\mathcal{P}}_T$ = PairwiseCompute(c, $\mathcal{F}$, C, $\overline{\mathcal{P}}$)
  end for
  Update $\mathcal{F}$, C with parameter average aggregation
  Update $\overline{\mathcal{P}}$ with Eq.5
end for
Fine-tune $\mathcal{F}$, C for $\mathcal{D}_T$

PairwiseCompute://the c-th source-target computation
for each local epoch t = 1 from 1 to X **do**
  Sample mini-batch from $\mathcal{D}_S^c$, $\mathcal{D}_T$
  Train the c-th source-target model with Eq.13, Eq.14
  Train the c-th discriminator $\mathcal{DI}_S^c$ with Eq. 11
  Compute local $\overline{\mathcal{P}}_S^k$, $\overline{\mathcal{P}}_T$ with Eq.3, Eq.4
  Update optimized $\mathcal{F}_S^c$, C_S^c, $\overline{\mathcal{P}}_S^c$, $\overline{\mathcal{P}}_T$
end for

8 EVALUATION

In this section, we evaluate the performance of our proposed FDAS intending to answer the following questions:

- Q1: Does FDAS outperform the baselines on the various human sensing datasets in the UFDA setting?
- Q2: How effective is each component in the design of our proposed FDAS?
- Q3: Has FDAS successfully learned features that can narrow the gap between the source and target domains?
- Q4: What is the wall-clock training time and computation/communication overhead of FDAS?

8.1 Experiment Methodology

Datasets. To demonstrate the generality of FDAS, we evaluate the performance of FDAS on four public human sensing datasets including WIFI dataset, Depth Image dataset, UWB dataset and a real-world smartphone dataset.

- **WIFI datasets:** Widar3.0 [53]is a public WIFI-based gesture recognition dataset, which is provided by researchers at Tsinghua University. The dataset is collected from 16 users (including both men and women). In this dataset, 16 users perform different gestures (including push & pull, sweep, clap, slide, draw a circle and draw zigzag) in three different environments, five different positions and five orientations.
- **Depth images datasets:** Depth image dataset [33] is a public image-based gesture recognition dataset using a depth camera. In this dataset, 9 users perform five types of gestures (including good, ok, victory, stop and fist) in three environments (outdoor, dark, and indoor).
- **IMU datasets:** IMU dataset [33] is a public IMU-based walking activity recognition dataset. In this dataset, 7 users conduct three walking activities (including walking in the corridor, walking upstairs and walking downstairs) in two buildings.
- **UWB datasets:** UWB datasets [33] is a public UWB-based human movement detection dataset. In this dataset, it detects whether someone (including 7 users) passes through in 3 different environments (including a parking lot, corridor and room). The objective of human movement detection is a binary classification task, aimed at determining whether a person is passing through a certain area or not.
- **Smartphone datasets:** Smartphone dataset is a real-world human activity recognition dataset, which is collected by us. We develop an android app to collect HAR data using users' own smartphone IMU module We invited each user to perform 5 activities, including walking, typing, phone, sitting and standing with different smartphones originated from distinct manufacturers. To preprocess the IMU data, we resample it to a frequency of 50Hz, and apply a sliding time window of 2 seconds. This results in the generation of a 750-dimensional feature for each data sample.

Baselines:

- **Source-only:** The source-only method combines source domains and trains a single model without domain adaptation. Thus, the source-only method can server as the lower-bound.
- **f-EI:** EI [17]is a state-of-the-art adversarial cross-domain method, which aligns sample features from source and target domains in feature space by domain adversarial method. EI need to access both source and target domain raw data. In our federated setting, we follow the original design to reimplement EI and only modify the input unit of the model to adapt to the input of different human sensing datasets. Moreover, we combined the FedAvg algorithm with EI to implement the federated version of EI, denoted by f-EI.
- **FADA:** FADA [37] is a federated domain adaptation method using domain adversarial and domain disentangling. FADA also utilize adversarial alignment to minimize the discrepancy between the target domain and the source domain and then enhances knowledge transfer through feature disentanglement. In our experiments, we also only modify the input unit of the model to adapt to the input of different human sensing datasets.

Table 3. The UFDA approaches on Widar3.0(BVP) (cross-user).

Method	user5	user10	user11	user12	user13	user14	user15	user16	avg
source-only	$73.57_{\pm 0.72}$	$80_{\pm 0.68}$	$71.85_{\pm 0.46}$	$61.33_{\pm 0.67}$	$78.86_{\pm 0.41}$	$78_{\pm 0.67}$	$80.19_{\pm 0.42}$	$73.59_{\pm 0.51}$	74.71
f-EI	$73.86_{\pm 0.48}$	$79.43_{\pm 0.69}$	$76.77_{\pm 0.57}$	$70.22_{\pm 1.02}$	$81.32_{\pm 0.68}$	$80.44_{\pm 1.01}$	$84_{\pm 0.85}$	$77.77_{\pm 0.4}$	77.98
FADA	$73.41_{\pm 1.09}$	$83.51_{\pm 0.72}$	$78.6_{\pm 0.67}$	$72.14_{\pm 0.62}$	$83.24_{\pm 0.25}$	$83.47_{\pm 0.59}$	$86.47_{\pm 0.59}$	$86.68_{\pm 0.72}$	80.93
KD3A	$79.87_{\pm 0.47}$	$85.22_{\pm 0.57}$	$82.04_{\pm 0.11}$	$73.73_{\pm 0.51}$	$88.94_{\pm 0.27}$	$82.45_{\pm 0.39}$	$85.86_{\pm 0.67}$	$89.22_{\pm 0.47}$	83.42
FDAS	$\mathbf{85.97_{\pm 0.61}}$	$\mathbf{90.06_{\pm 0.46}}$	$\mathbf{83.68_{\pm 0.43}}$	$\mathbf{78.63_{\pm 0.35}}$	$\mathbf{92.49_{\pm 0.43}}$	$\mathbf{86.2_{\pm 0.33}}$	$\mathbf{88.96_{\pm 0.24}}$	$\mathbf{92.34_{\pm 0.57}}$	**87.29**

- **KD3A:** KD3A [10] is a knowledge distillation method. KD3A proposes a multi-source model knowledge distillation method based on Knowledge Vote and BatchNorm MMD to obtain high-quality target model. In our experiments, we also only modify the input unit of the model to adapt to the input of different human sensing datasets. KD3A deviates from the federated learning paradigm and aims to achieve a high-performance target model rather than a viable global model.

Experiment Testbed: In order to conduct a comprehensive evaluation of a federated learning system, it is crucial to have an experimental testbed that accurately simulates the characteristics of real-world federated learning scenarios. Thus, we consider the heterogeneous devices and networks of the clients. To be specific, we utilized one Jetson Xavier NX device, one Jetson Nano device, and six Raspberry Pi 4B+ devices as edge clients, while a 4-core 2G cloud servers a trusted central cloud for federated aggregation. We place the heterogeneous IoT devices in different locations (room1, room2 and room3) to achieve distinct network environments. All the devices connect to the server through a `frp` [1] proxy. In the Widar3.0, UWB and Imu datasets, sensor data collected from the same individual is considered private and each individual (domain) is viewed as a client. In the depth images datasets, sensor data collected within the same place (domain) is considered private and each place is viewed as a client. In the smartphone datasets, data collected by the same sensor is considered private and each smartphone (domain) is viewed as a client.We use 90 percent of data for training and 10 percent for testing, respectively. We perform each experiment 10 times and provide 95% confidence intervals.

8.2 Overall Performance

In this section, we present our results comparing FDAS against other state-of-art methods on cross-user, cross-place and cross-device scenarios in the UFDA setting.

8.2.1 The performance on cross-user datasets. We evaluate the cross-user performance of FDAS on Widar3.0, UWB and IMU datasets. Note that we use BVP features of Widar3.0 datasets, which have been demonstrated to mitigate the impact of different places. However, the BVP features still exist a significant domain-shift issue among different users because of the behavior differences between human subjects. Thus, the detailed experiment setup is as follows: (1) We consider the BVP data from all 8 users in Widar3.0 dataset who performed the same set of actions as clients. We designate one of these clients as the target domain sequentially, while the remaining clients were treated as the source domains. The experiment results as shown in Table 3. (2) We divided all data into four parts, each of which includes data from different places to reduce the impact of cross-place. The detailed experiment setup and results are presented in Table 4 and 5 (UWB:user#= user# + $\frac{1}{4}$user5,6,7 and IMU:user#=user# + $\frac{1}{4}$user4,5,6). Our feature extractor consists of two convolutional layers and three fully connected layers. We use a fully connected layer as the classifier and two fully connected layers as the discriminator. Our proposed FDAS demonstrates average accuracy rates of 87.29%, 88.03%, and 68.63%, respectively, surpassing the performance of all other baseline approaches.

8.2.2 The performance on cross-place datasets. We evaluate the cross-place performance of FDAS on depth images datasets. We merge the data from the same places across different users and grouped them into three

Method	user4	user3	user2	user1	avg
source-only	$64.33_{\pm 0.62}$	$75.65_{\pm 1.03}$	$70.58_{\pm 0.79}$	$70.42_{\pm 0.53}$	70.24
f-EI	$68.26_{\pm 0.47}$	$80.73_{\pm 0.58}$	$72.31_{\pm 0.95}$	$72.48_{\pm 0.66}$	73.44
FADA	$70.36_{\pm 0.42}$	$88.64_{\pm 0.68}$	$82.35_{\pm 0.83}$	$77.3_{\pm 1.03}$	79.66
KD3A	$\mathbf{71.42_{\pm 0.55}}$	$94.03_{\pm 1.52}$	$87.88_{\pm 0.76}$	$94.14_{\pm 0.91}$	86.87
FDAS	$70.53_{\pm 0.92}$	$\mathbf{95.32_{\pm 0.65}}$	$\mathbf{89.59_{\pm 0.61}}$	$\mathbf{96.97_{\pm 0.79}}$	**88.03**

Table 4. The UFDA approaches on UWB datasets (cross-user).

Method	user3	user2	user1	use0	avg
source-only	$66.52_{\pm 0.63}$	$56.33_{\pm 0.59}$	$43.27_{\pm 0.53}$	$51.39_{\pm 0.71}$	54.38
f-EI	$72.51_{\pm 0.73}$	$64.46_{\pm 0.57}$	$49.37_{\pm 0.44}$	$52.91_{\pm 0.65}$	59.81
FADA	$73.83_{\pm 0.26}$	$61.64_{\pm 0.78}$	$56.58_{\pm 0.37}$	$53.11_{\pm 1.12}$	61.29
KD3A	$71.52_{\pm 0.37}$	$\mathbf{68.09_{\pm 0.92}}$	$52.89_{\pm 1.07}$	$66.77_{\pm 0.91}$	64.82
FDAS	$\mathbf{76.38_{\pm 0.76}}$	$66.98_{\pm 0.75}$	$\mathbf{61.67_{\pm 0.81}}$	$\mathbf{69.49_{\pm 0.32}}$	**68.63**

Table 5. The UFDA approaches on IMU datasets (cross-user).

Method	P1	P2	P3	avg
source-only	$61.41_{\pm 0.61}$	$71.3_{\pm 0.34}$	$72.59_{\pm 0.87}$	68.43
f-EI	$64.78_{\pm 0.4}$	$75.96_{\pm 0.7}$	$74.9_{\pm 0.76}$	71.88
FADA	$66.05_{\pm 1.09}$	$77.69_{\pm 0.87}$	$75.43_{\pm 0.44}$	73.05
KD3A	$64.91_{\pm 0.64}$	$79.34_{\pm 0.42}$	$77.32_{\pm 0.45}$	73.86
FDAS	$\mathbf{79.98_{\pm 0.63}}$	$\mathbf{82.02_{\pm 0.59}}$	$\mathbf{78.8_{\pm 0.67}}$	**80.2**

Table 6. The UFDA approaches on depth images datasets (cross-place).

Method	D1	D2	D3	D4	avg
source-only	$83.62_{\pm 1.04}$	$78.64_{\pm 0.77}$	$71.29_{\pm 0.8}$	$90.82_{\pm 0.17}$	81.09
f-EI	$90.17_{\pm 0.54}$	$86.83_{\pm 0.59}$	$89.09_{\pm 0.22}$	$90.95_{\pm 0.21}$	89.26
FADA	$90.68_{\pm 0.44}$	$85.42_{\pm 0.32}$	$87.11_{\pm 0.12}$	$93.1_{\pm 0.52}$	89.08
KD3A	$91.59_{\pm 0.34}$	$85.26_{\pm 0.88}$	$\mathbf{94.37_{\pm 0.34}}$	$91.93_{\pm 0.45}$	90.79
FDAS	$\mathbf{94.48_{\pm 0.58}}$	$\mathbf{92.91_{\pm 0.48}}$	$93.6_{\pm 0.56}$	$\mathbf{96.18_{\pm 0.65}}$	**94.29**

Table 7. The UFDA approaches on smartphone datasets (cross-device).

clients, (1) P1: outdoor-normal, (2) P2: indoor-normal, (3) P3: indor-dark. We select one of these clients as the target domain sequentially, while the remaining clients were treated as the source domains. We use ResNet-101 [14] as the backbone of the feature extractor. We use a fully connected layer as the classifier and two fully connected layers as the discriminator. The experiment results on Depth images Datasets as shown in Table 6. Our proposed FDAS achieves 80.2% and outperforms all other baselines.

8.2.3 The performance on cross-device datasets. We evaluate the cross-place performance of FDAS on smartphone datasets. We merged the data from the same users across different devices and grouped them into four clients (D1, D2, D3 and D4). We also use a fully connected layer as the classifier and two fully connected layers as the discriminator. The experimental results on IMU Datasets as shown in Table 7. Our proposed FDAS achieves 94.29% and outperforms all other baselines.

8.2.4 The performance in real-world UFDA scenarios. We next discuss the performance of FDAS in real-world ubiquitous computing scenarios. The concept of domain describes data distribution and data from different domains often exhibit heterogeneity, while the concept of client emphasizes the scope of privacy data protection. Therefore, there are three distinct types of relationships between domains and clients in real-world scenarios.

- **Scenario 1: each domain is treated as an independent client.** The client is defined to protect the data from each domain. For example, the Human Activity Recognition (HAR) data collected from each user constitutes a domain, and the data of each user is private. Thus, exchanging raw data between each domain is prohibited in the scenario.
- **Scenario 2: each client contains multiple domains.** The client is defined to protect the data from multiple domains. For example, The HAR data collected from different users represents different domains, but privacy concerns only arise when the data comes from different rooms (clients). Thus, cross-domain data exchange within the same client is allowed, while cross-client data exchange is prohibited in this scenario.
- **Scenario 3: each domain contains multiple clients.** The client is defined to protect the data from a certain part of a domain. For example, The HAR data collected from different users represents different domains, but the privacy concerns pertain to data collected from different data collection devices (clients). Thus, data exchange between different clients within the same domain is prohibited.

Table 8. Data partition for the three scenarios, where the number of user#_1 = user#_2 and user# = user#_1 + user#_2. Each cell in the table represents a distinct client, and each unique user# corresponds to a different domain.

scenario	partitioned source domains							
1	user5	user10	user11	user12	user13	user14	user15	user16
2	user5 ∪ user10		user11 ∪ user12		user13 ∪ user14		user15 ∪ user16	
3	user5_1	user10_1	user11_1	user12_1	user13_1	user14_1	user15_1	user16_1
	user5_2	user10_2	user11_2	user12_2	user13_2	user14_2	user15_2	user16_2

FDAS has shown the effectiveness in the scenario 1. We further evaluate the performance in the scenario 2 & 3 on widar3.0 datasets by data partitioning. The detailed data partition is presented in Table 8. We consider one user (domain) as the target domain, and the other users not belonging to the same target client are considered as the source domains. The source domains can be partitioned into different clients in a coarse-grained or fine-grained manner. We report the results of our FDAS performance in the three scenarios. We also provide the performance of FDAS with centralized training, when all the source domains are considered as a client. In the centralized training manner, our method only requires the pre-mixing of all the source data beforehand. The feature selection and refinement mechanism still works. The results are presented in Fig 4.

We can have the following observations: (1) The performance of FDAS in scenario 2 is slightly better than in scenario 1. This is because each client has more training data and enables the local model to learn the more generic feature of the classes in scenario 2. Thus, the global model may have more opportunities to learn domain-invariant features from pairwise domain-level alignment. (2) The performance of FDAS in scenario 3 is slightly lower than in scenario 1. This discrepancy arises from the necessity to partition the target domain into two clients, whereby we are constrained to utilize only one of them as the target training data due to privacy concerns. The limited training data leads to a degradation of feature generality in the specific classes learned during the pairwise source-target computation. We believe that increased local training data would ameliorate the performance degradation incurred by this dataset partition. (3) FDAS (centralized) can achieve better performance than scenarios with strict privacy constraints, where data sharing among source domains is prohibited. This is attributed to the fact that FDAS with centralized training alleviates negative transfer (see Section 1) problem caused by data heterogeneity among source domains.

Thus, scenarios with higher privacy requirements may lead to relatively lower cross-domain accuracy of the federated global model. However, stringent privacy settings are prevalent in many scenarios, such as wearable devices. Our method demonstrates the ability to achieve performance closest to those with no privacy requirements under highly strict privacy constraints in the real-world ubiquitous computing scenarios.

8.2.5 The performance on the target domain with limited labeled data. We further consider the additional situation in real-world scenarios, where the target domain has a limited number of labels. Our method is also suitable for this scenario. We utilize the true label to calibrate the pseudo labels generated during our training process. The pseudo labels are used as conditional information for our adversarial training and similarity-aware fine-tuning. We conduct the experiments on the WiFi dataset to show the performance of our method. We use user11, user12, and user14 data as an example because these data exhibit the lower cross-domain accuracy as the target domain. We measured the performance of our method when these users have 5%, 10%, 15%, 20%, and 25% labels, respectively. The results are shown in Fig. 5.

We note an increasing performance trend in our method with the growing number of labels. This is primarily attributed to two factors. Firstly, the use of labels as correct conditional information guides adversarial domain alignment, thereby further alleviating the problem of mode collapse during adversarial training. Secondly, the limited labeling is applied to enhance our similarity-aware fine-tuning strategy.

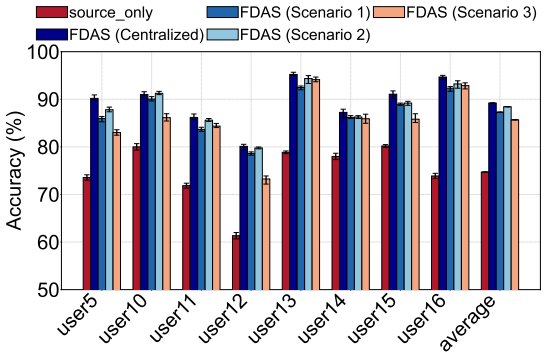


Fig. 4. The performance of FDAS in the three real-word scenarios.

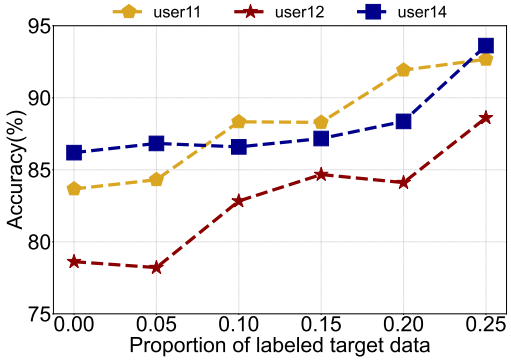


Fig. 5. The performance of FDAS when the target domain has limited labeled data.

8.3 Analysis of FDAS

In this section, we present ablation studies and a detailed analysis of our proposed FDAS. Our primary focus lies in proving both semantic-level and domain-level alignment. We also provide the analysis of the key component feature selection mechanism, while the other component studies are presented in Appendix A.

8.3.1 Semantic-level and domain-level alignment analysis. We implement FDAS without semantic-level alignment (SGCC) and FDAS without domain-level alignment (SRAC) to analyze the contributions of the two different alignments. We conduct the experiments on widar3.0 datasets. The result is shown in Fig. 6. We observe that the combination of semantic-level and domain-level alignment achieves the highest model performance. The misalignment of environment-dependent features in the single domain-level alignment limits the upper bound of the performance of the adversarial strategy in the UFDA setting. The discrepancy between the source domains and target domain results in limited efficacy of a single semantic-level alignment approach due to the lack of high-quality cross-domain semantic knowledge.

The feature selection mechanism is the key component of the semantic-level and domain-level alignment. We have discussed the difference between feature selection and client selection in section 3.3. Feature selection can reduce negative transfer while minimizing the need to discard source data, while client selection will discard the whole source domain. To validate the efficiency of the feature selection mechanism, we track the selected features and record the number of transferred bytes in each pairwise source-target computation at different stages of the training process (see Fig. 7). We have the following observations: (1) The transferred bytes count of our proposed FDAS is consistently lower than f-EI and FADA for each pairwise source-target computation. (2) The number of transferred bytes for most source-target pairwise computations (especially for user4-target and user7-target pairwise) generally follows a pattern of initial increase, subsequent decrease, and then increase again. This is because that our feature selection mechanism increases the proportion of selected source features with high similarity to the target domain and reduces the proportion of selected source features with low similarity to the target domain. It verifies that our feature selection mechanism can select the features that are easy to transfer first, and then use the learned knowledge to transfer more hard-to-transfer features. In this way, FDAS can improve the accuracy and reduce the network footprint compared to f-EI and FADA, while f-EI and FADA always need to transfer the whole features.

8.3.2 The learned feature representations analysis. The learned feature representations are the features extracted by the converged federated model on each domain. The feature can be used to evaluate the cross-domain

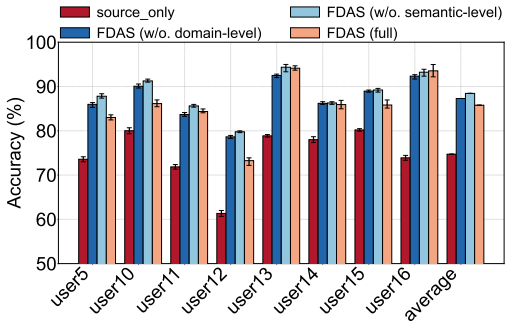


Fig. 6. The ablation study of semantic-level and domain-level alignment.

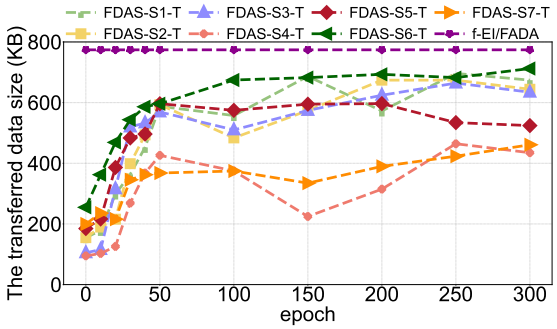


Fig. 7. The transferred number of bytes for each source-target computation. FDAS-S#-T represents FDAS-user#-T. The target domain is user16 and the remaining 7 users are treated as the source domains.

performance of the federated model. To qualitatively evaluate the learned representation and further verify the effectiveness of FDAS, we provide the t-SNE embeddings of the feature representation on both the source domain and target domain, which is shown in Fig 8 and 9. In Fig 8, it is the feature extracted by FADA method. We can observe that the data samples from the source domains are divided into 5 clusters. However, there still exists a large variance between the features from the target domain (green points) and the source domains (other color points). In Fig. 9, it is the feature extracted by our proposed FDAS method. Both the data samples from the target (red points) and source (other color points) domains are clearly divided into 5 clusters. The feature visualization illustrates that our proposed FDAS can minimize the discrepancy between the source and target domains in the UFDA setting.

To better analyze the effectiveness of the learned feature representations, we utilize $\mathcal{A}$-distance [2] to evaluate feature discrepancy between the source domains and target domain. We calculate the approximate A-distance $\hat{d} = 2(1 - 2\epsilon)$ [25] for each pair-wise source-target clients on depth images. ϵ is the generalization error of a two-sample classifier (e.g. kernel SVM), which tries to distinguish the source features and target features. The results are shown in Fig. 10. We observe that the $\hat{d}$ on our proposed FDAS are smaller than source-only features, f-EI and FADA features. It can demonstrate that the extracted source-target features by FDAS are harder to be distinguished and validate the reduction of descrepancy between the source and target domains.

8.4 The Overhead Analysis

In this section, we provide the wall-clock comparisons between source-only, f-EI, FADA and our proposed FDAS to report the wall-clock training time and computation/communication overhead. Note that KD3A deviates from the federated learning paradigm, and fails to attain a viable global model (only the target model is achieved). KD3A does not incorporate federated training, thereby sacrificing the federated training efficiency. Thus, we do not analyze the client-side overhead comparison between our proposed FDAS and KD3A for the sake of fairness. We conducted experiments on Widar3.0 datasets, with user13 client data designated as the target domain, and the data from the clients of the remaining users as the source domain. We treat each domain as a separate client. Extensive experiments demonstrate that FDAS can substantially improve cross-domain performance while incurring relatively low computation and communication overhead.

8.4.1 The wall-clock training time. We report the wall-clock time to reach the final accuracy for FDAS and other methods, which are shown in Fig. 11. It can be observed that FDAS achieves a notable improvement on

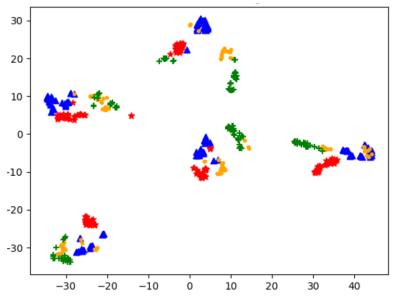


Fig. 8. FADA: Feature visualization on the Depth images dataset by t-SNE.

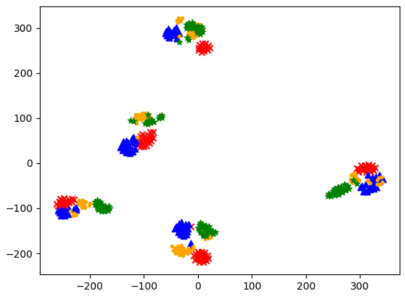


Fig. 9. FDAS: Feature visualization on the Depth images dataset by t-SNE.

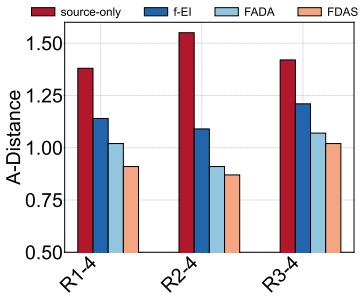


Fig. 10. The $\mathcal{A}$-distance of source-only, f-EI, FADA and FDAS on the Depth images dataset.

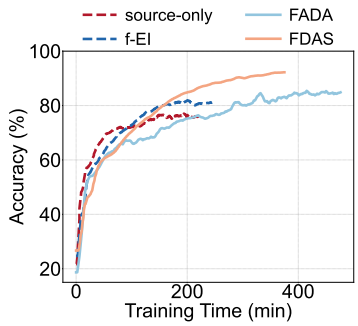


Fig. 11. The wall-clock training time

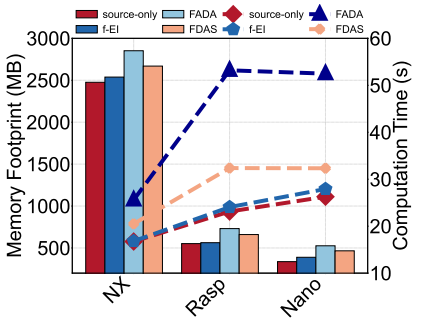


Fig. 12. The computation overhead

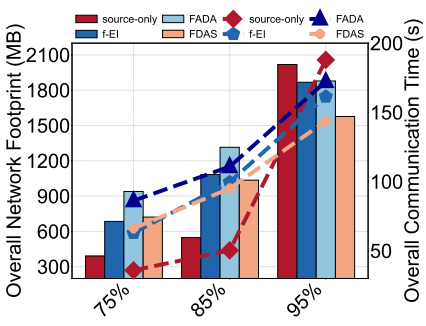


Fig. 13. The communication overhead

wall-clock time consumption compared to FADA, and a slight degradation compared to source-only and f-EI. The reasons are as follows: (1) FDAS mainly consists of an additional adversarial network training and prototype computation, compared to the source-only method. (2) FADA need to train two additional adversarial network compared to the source-only method, and the adversarial network training is more time-consuming than our prototype computation (GPUs are not easily available on ubiquitous computing devices.). (3) f-EI solely entails supplementary adversarial network training in comparison to the source-only method, yet culminates in a relatively diminished final accuracy.

8.4.2 The computation overhead. To illustrate the impact of heterogeneous devices on FDAS computation overhead, we measure memory usage and computation time of three different heterogeneous devices (Jetson Nano, Raspberry Pi, and Jetson NX) for each round. Among these heterogeneous devices, Jetson NX, with GPU acceleration enabled, occupies a relatively larger amount of memory. The results are shown in Fig. 12. We observe that the memory usage and computation time of FDAS are significantly lower on different devices compared to FADA, while slightly higher than the source-only and f-EI methods. Fig. 12 also provides evidence that FDAS demands fewer GPU resources on the client side in comparison to FADA, rendering it better suited for ubiquitous computing scenarios.

8.4.3 The communication overhead. We measure the overall network footprint and computation time on heterogeneous networks for FDAS and other methods. We report the communication network usage and communication

time for all methods at the final accuracy thresholds of 75%, 85%, and 95%, which is shown in Fig. 13. Reaching a specific accuracy level is contingent on a combination of communication rounds, data transmission volume, and network bandwidth. We have the following two observations: (1) In the initial training phases, FDAS exhibits lower communication overhead compared to FADA and f-EI, while surpassing that of the source-only method. This is because FDAS with the federated feature selection mechanism transmits fewer features compared to FADA and f-EI, while conveying more of the selected features in contrast to the source-only method. (2) In the later stages of training, the overall communication cost of FDAS becomes lower than that of source-only and all other methods. This is because FDAS, with its two-level alignments, can align the semantic information of data from multiple clients, thereby mitigating the data heterogeneity among clients and allowing the model to converge in fewer communication rounds compared to the source-only method.

9 DISCUSSION

9.1 The Effect of Heterogeneous Devices and Networks

The introduction of heterogeneous computation and communication capabilities has increased the wall-clock training time, while the accuracy of our proposed FDAS has remained unaffected. This is because the presence of stragglers, characterized by low computation capabilities or limited communication bandwidth. Our FDAS is established on FedAvg (the classic synchronous FL framework). In the classic FL framework, relatively higher computation power or communication bandwidth clients have to wait for the completion of training on the stragglers. The waiting time is the primary cause for the increased training time.

Existing work has explored many effective methods to mitigate the straggler problem. (1) One promising method is to design an efficient node selection algorithm, such as TiFL [3] and ClusterFL [33]. The fundamental concept revolves around the adaptive selection of clients with similar training times to participate in the training process during each training round. (2) Another viable approach is to offer an asynchronous federated training method, such as HFL [22]. The framework comprises a synchronous updater and an asynchronous updater. The asynchronous updater integrates model updates from stragglers into the federated model training process, albeit with a slight delay compared to the synchronous updater. We plan to integrate an effective method for mitigating stragglers into FDAS as our future work to reduce the impact of heterogeneous devices on the training efficiency of FDAS.

9.2 UFDA and Data Heterogeneous under FL

UFDA emphasis on the target domain having a limited number of labeled samples or being unlabeled, whereas heterogeneity under FL typically involves scenarios with an abundance of labeled data. Our research problem UFDA, is dedicated to the task of training a universal model among multiple source domains under FL, with the primary objective of enabling seamless adaptation to an unlabeled target domain. The different domains refer to data collected from different users or environments, typically characterized by data heterogeneity. Hence, the target domain in the UFDA setting possesses two distinctive characteristics: (1) The target domain is heterogeneous compared to all source domains. (2) The target domain lacks labels. Thus, the UFDA problem is a domain adaptation problem in the context of federated learning, which differs from data heterogeneity under FL due to the absence of the labeled target domain.

To address the UFDA problem, we need to reduce data heterogeneity in the latent space between the target and source domains through domain adaptation. (see Fig. 1). Moreover, if the target domain has a sufficient number of labels, the UFDA problem will degrade into data heterogeneity under FL. Thus, the UFDA and data heterogeneity under FL problem are complementary.

10 CONCLUSION

We propose an effective approach FDAS to address the UFDA problem. We explore the significant barriers when employing the adversarial strategy in the UFDA setting. The main idea of FDAS is to perform semantic-level alignment and domain-level alignment simultaneously for each source-target pair computation. Thus, we can improve the semantic quality of the aligned features thereby reducing the misalignment of environment-dependent features. In addition, we devise federated feature selection and feature refinement mechanisms to enhance the domain-level and semantic-level alignment. Extensive experiments demonstrate the superior effectiveness and generalization of FDAS.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China under Grant No. 62272407 and 62072396, the "Pioneer" and "Leading Goose" R&D Program of Zhejiang under grant No. 2023C01033, and the National Youth Talent Support Program. The work is also supported by Information Technology Center and State Key Lab of CAD&CG, ZheJiang University.

REFERENCES

[1] 2023. frp. https://github.com/fatedier/frp.
[2] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. 2010. A theory of learning from different domains. *Machine learning* 79 (2010), 151–175.
[3] Zheng Chai, Ahsan Ali, Syed Zawad, Stacey Truex, Ali Anwar, Nathalie Baracaldo, Yi Zhou, Heiko Ludwig, Feng Yan, and Yue Cheng. 2020. Tifl: A tier-based federated learning system. In *Proceedings of the 29th international symposium on high-performance parallel and distributed computing*. 125–136.
[4] Chaoqi Chen, Weiping Xie, Wenbing Huang, Yu Rong, Xinghao Ding, Yue Huang, Tingyang Xu, and Junzhou Huang. 2019. Progressive feature alignment for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 627–636.
[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*. PMLR, 1597–1607.
[6] Hyunsung Cho, Akhil Mathur, and Fahim Kawsar. 2022. Flame: Federated learning across multi-device environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–29.
[7] Bo Dai, Sanja Fidler, Raquel Urtasun, and Dahua Lin. 2017. Towards diverse and natural image descriptions via a conditional gan. In *Proceedings of the IEEE international conference on computer vision*. 2970–2979.
[8] Emily L Denton et al. 2017. Unsupervised learning of disentangled representations from video. *Advances in neural information processing systems* 30 (2017).
[9] Chao Feng, Nan Wang, Yicheng Jiang, Xia Zheng, Kang Li, Zheng Wang, and Xiaojiang Chen. 2022. Wi-Learner: Towards One-shot Learning for Cross-Domain Wi-Fi based Gesture Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–27.
[10] Haozhe Feng, Zhaoyang You, Minghao Chen, Tianye Zhang, Minfeng Zhu, Fei Wu, Chao Wu, and Wei Chen. 2021. KD3A: Unsupervised Multi-Source Decentralized Domain Adaptation via Knowledge Distillation.. In *ICML*. 3274–3283.
[11] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The journal of machine learning research* 17, 1 (2016), 2096–2030.
[12] Ruiyang Gao, Mi Zhang, Jie Zhang, Yang Li, Enze Yi, Dan Wu, Leye Wang, and Daqing Zhang. 2021. Towards position-independent sensing for gesture recognition with Wi-Fi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–28.
[13] Yves Grandvalet and Yoshua Bengio. 2004. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems* 17 (2004).
[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
[15] Liang Hou, Qi Cao, Huawei Shen, Siyuan Pan, Xiaoshuang Li, and Xueqi Cheng. 2022. Conditional gans with auxiliary discriminative classifier. In *International Conference on Machine Learning*. PMLR, 8888–8902.
[16] Yash Jain, Chi Ian Tang, Chulhong Min, Fahim Kawsar, and Akhil Mathur. 2022. ColloSSL: Collaborative self-supervised learning for human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–28.

[17] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. 2018. Towards environment independent device free human activity recognition. In *Proceedings of the 24th annual international conference on mobile computing and networking*. 289–304.

[18] Chenning Li, Xiao Zeng, Mi Zhang, and Zhichao Cao. 2022. PyramidFL: A fine-grained client selection framework for efficient federated learning. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 158–171.

[19] Jingjing Li, Erpeng Chen, Zhengming Ding, Lei Zhu, Ke Lu, and Zi Huang. 2019. Cycle-consistent conditional adversarial transfer networks. In *Proceedings of the 27th ACM international conference on multimedia*. 747–755.

[20] Junnan Li, Pan Zhou, Caiming Xiong, and Steven CH Hoi. 2020. Prototypical contrastive learning of unsupervised representations. *arXiv preprint arXiv:2005.04966* (2020).

[21] Wei-Hong Li, Xialei Liu, and Hakan Bilen. 2022. Cross-domain few-shot learning with task-specific adapters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7161–7170.

[22] Xingyu Li, Zhe Qu, Bo Tang, and Zhuo Lu. 2021. Stragglers are not disaster: A hybrid federated learning algorithm with delayed gradients. *arXiv preprint arXiv:2102.06329* (2021).

[23] Yitong Li, Zhe Gan, Yelong Shen, Jingjing Liu, Yu Cheng, Yuexin Wu, Lawrence Carin, David Carlson, and Jianfeng Gao. 2019. Storygan: A sequential conditional gan for story visualization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6329–6338.

[24] Yadong Li, Dongheng Zhang, Jinbo Chen, Jinwei Wan, Dong Zhang, Yang Hu, Qibin Sun, and Yan Chen. 2022. Towards domain-independent and real-time gesture recognition using mmwave signal. *IEEE Transactions on Mobile Computing* (2022).

[25] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. 2015. Learning transferable features with deep adaptation networks. In *International conference on machine learning*. PMLR, 97–105.

[26] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. 2018. Conditional adversarial domain adaptation. *Advances in neural information processing systems* 31 (2018).

[27] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.

[28] Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. 2016. Unrolled generative adversarial networks. *arXiv preprint arXiv:1611.02163* (2016).

[29] Umberto Michieli and Mete Ozay. 2021. Prototype guided federated learning of visual feature representations. *arXiv preprint arXiv:2105.08982* (2021).

[30] Mehdi Mirza and Simon Osindero. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014).

[31] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. 2017. Unified deep supervised domain adaptation and generalization. In *Proceedings of the IEEE international conference on computer vision*. 5715–5725.

[32] Dinh C Nguyen, Quoc-Viet Pham, Pubudu N Pathirana, Ming Ding, Aruna Seneviratne, Zihuai Lin, Octavia Dobre, and Won-Joo Hwang. 2022. Federated learning for smart healthcare: A survey. *ACM Computing Surveys (CSUR)* 55, 3 (2022), 1–37.

[33] Xiaomin Ouyang, Zhiyuan Xie, Jiayu Zhou, Jianwei Huang, and Guoliang Xing. 2021. Clusterfl: a similarity-aware federated learning system for human activity recognition. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. 54–66.

[34] Brooks Paige, Jan-Willem van de Meent, Alban Desmaison, Noah Goodman, Pushmeet Kohli, Frank Wood, Philip Torr, et al. 2017. Learning disentangled representations with semi-supervised deep generative models. *Advances in neural information processing systems* 30 (2017).

[35] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2010), 1345–1359.

[36] Xingchao Peng, Zijun Huang, Ximeng Sun, and Kate Saenko. 2019. Domain agnostic learning with disentangled representations. In *International Conference on Machine Learning*. PMLR, 5102–5112.

[37] Xingchao Peng, Zijun Huang, Yizhe Zhu, and Kate Saenko. 2019. Federated adversarial domain adaptation. *arXiv preprint arXiv:1911.02054* (2019).

[38] Ariadna Quattoni, Michael Collins, and Trevor Darrell. 2008. Transfer learning for image classification with sparse prototype representations. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1–8.

[39] Konstantin Sozinov, Vladimir Vlassov, and Sarunas Girdzijauskas. 2018. Human activity recognition using federated learning. In *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)*. IEEE, 1103–1111.

[40] Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. 2021. Fedproto: Federated prototype learning over heterogeneous devices. *arXiv preprint arXiv:2105.00243* (2021).

[41] Yue Tan, Guodong Long, Jie Ma, Lu Liu, Tianyi Zhou, and Jing Jiang. 2022. Federated learning from pre-trained models: A contrastive learning approach. *arXiv preprint arXiv:2209.10083* (2022).

[42] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. 2019. Panet: Few-shot image semantic segmentation with prototype alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9197–9206.
[43] Ke Wang, Jiayong Liu, and Jing-Yan Wang. 2019. Learning domain-independent deep representations by mutual information minimization. *Computational Intelligence and Neuroscience* 2019 (2019).
[44] Zhengjie Wang, Zehua Huang, Chengming Zhang, Wenwen Dou, Yinjing Guo, and Da Chen. 2021. CSI-based human sensing using model-based approaches: a survey. *Journal of Computational Design and Engineering* 8, 2 (2021), 510–523.
[45] Xiaofu Wu, Suofei Zhang, Quan Zhou, Zhen Yang, Chunming Zhao, and Longin Jan Latecki. 2021. Entropy minimization versus diversity maximization for domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems* (2021).
[46] Minghao Xu, Hang Wang, Bingbing Ni, Qi Tian, and Wenjun Zhang. 2020. Cross-domain detection via graph-induced prototype alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12355–12364.
[47] Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu. 2018. Robust classification with convolutional prototype learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3474–3482.
[48] Jianfei Yang, Xinyan Chen, Dazhuo Wang, Han Zou, Chris Xiaoxuan Lu, Sumei Sun, and Lihua Xie. 2022. Deep learning and its applications to WiFi human sensing: A benchmark and a tutorial. *arXiv preprint arXiv:2207.07859* (2022).
[49] Mang Ye, Xu Zhang, Pong C Yuen, and Shih-Fu Chang. 2019. Unsupervised embedding learning via invariant and spreading instance feature. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6210–6219.
[50] Xu Yu, Dingjia Zhan, Lei Liu, Hongwu Lv, Lingwei Xu, and Junwei Du. 2021. A privacy-preserving cross-domain healthcare wearables recommendation algorithm based on domain-dependent and domain-independent feature fusion. *IEEE Journal of Biomedical and Health Informatics* 26, 5 (2021), 1928–1936.
[51] Jie Zhang, Zhanyong Tang, Meng Li, Dingyi Fang, Petteri Nurmi, and Zheng Wang. 2018. CrossSense: Towards cross-site and large-scale WiFi sensing. In *Proceedings of the 24th annual international conference on mobile computing and networking*. 305–320.
[52] Xie Zhang, Chengpei Tang, Kang Yin, and Qingqian Ni. 2021. WiFi-Based Cross-Domain Gesture Recognition via Modified Prototypical Networks. *IEEE Internet of Things Journal* 9, 11 (2021), 8584–8596.
[53] Yi Zhang, Yue Zheng, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2021. Widar3. 0: Zero-effort cross-domain gesture recognition with wi-fi. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

APPENDIX A

A THE DETAILED ABLATION STUDIES.

We provide detailed ablation studies for FDAS. We conduct evaluation experiments on Widar3.0 datasets to verify the effectiveness of feature selection mechanism, feature refinement mechanism and model fine-tuning strategy.

A.1 Feature Selection mechanism

The feature selection mechanism is shared by semantic-level alignment and domain-level alignment, while the refinement mechanism is only designed for domain-level alignment. Feature selection for SGCC is mainly expected to improve the global prototype and reduce contrastive misalignment. Feature selection for SRAC is mainly expected to reduce the bias accumulation caused by incorrect conditional information and alleviating negative transfer problems.

To investigate the effects of the feature selection mechanism on both domain-level alignment and semantic alignment, we implemented multiple ablations of FDAS: (1) version I: FDAS without feature selection mechanism both in SGCC and SRAC. (2) version II: FDAS without feature selection mechanism in SRAC. (3) version III: FDAS without feature selection mechanism in SGCC. We conduct the experiments on widar3.0 datasets and the result is shown in Fig. 14. We can observe that all the ablations have degraded performance. version III can demonstrate the crucial role of the feature selection mechanism in SGCC. This is due to the lack of a feature selection mechanism that will lead to low-quality global prototype, and SGCC will cause a large number of contrastive misalignments under the guidance of low-quality global prototype. version II can prove the important role of feature selection in SRAC. This is because the feature selection mechanism effectively excludes the alignment of irrelevant features at the feature level to alleviate negative transfer, and reduces the bias accumulation caused by incorrect conditional information. The fact that version II sometimes performs worse than version I implies the significant impact of negative transfer on SRAC. Without a feature selection mechanism, SRAC may suffer from negative transfer, leading to a performance decrease of the federated global model, and even lower performance than the source-only method.

A.2 Feature refinement mechanism

We implement FDAS without feature refinement mechanism to evaluate the effectiveness of feature refinement mechanism. The result is shown in Fig. 15. We observe that the performance of FDAS without the feature

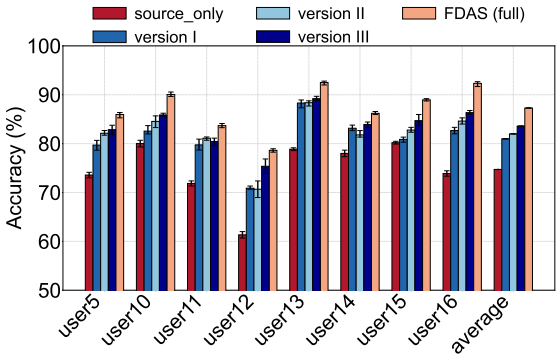


Fig. 14. The ablation study of feature selection mechanism.

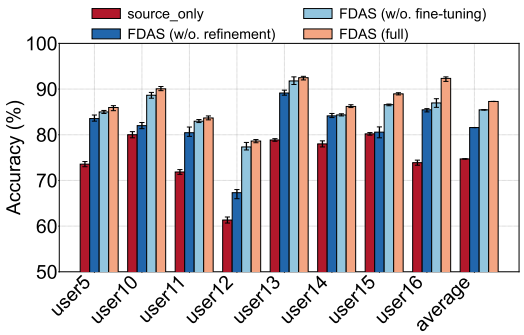


Fig. 15. The ablation study of feature refinement and fine-tuning.

refinement mechanism is inferior compared to FDAS. This is because the refined features contain more semantic information compared to raw features, which can enhance similarity within the same class and reduce similarity between different classes. In this way, the feature refinement mechanism can alleviate the misalignment of pairwise source-target features and improve the performance.

A.3 A similarity-aware model fine-tuning strategy

Our similarity-aware model fine-tuning strategy aims to generate high-quality pseudo-labels by measuring the similarity between target features and the domain-invariant prototypes generated by SGCC and SRAC. The result is shown in Fig. 15. We can observe that FDAS with fine-tuning outperforms FDAS without fine-tuning. This is due to the fact that personalizing fine-tuning on the final federated global model facilitates the extraction of domain-specific features, thereby can achieve a further performance improvement.

APPENDIX B

B SENSITIVITY OF HYPER-PARAMETERS

We conduct experiments on Widar3.0 datasets to analyze the sensitivity of hyper-parameter τ, the combination of α_s and β_s, and the combination of α_t and β_t in our approach. The parameter τ controls the proportion of local features that are selected, where only local features with similarity to the global prototype above τ are selected. We adjusted τ over the range of [0, 0.95] (see Fig. 16) and found that a smaller τ led to lower accuracy, indicating a higher likelihood of negative transfer. Conversely, when τ was too large, there was a significant decrease in accuracy, indicating that too few features were selected and preventing the model from improving. Therefore, we considered gradually increasing τ with epoch to avoid situations where τ is either too small or too large, and we achieved higher accuracy (88.67%) compared to using a fixed τ (86.32%).

The combination of α_s and β_s is used to modulate the contributions of adversarial alignment and semantic alignment in the source domain. We conducted a parameter search over the ranges of [0, 0.95] for τ and [0, 0.16] for β_s, and the corresponding results are depicted in Fig. 17. We observed that higher accuracy was achieved when α_s was within the interval [0.75, 0.85] and β_s was within [0.04, 0.12]. We also conduct experiments on the combination of α_t and β_t for the target domain. The intervals for α_t and β_t were both set to [0.75-0.95]. As illustrated in Fig. 18, we observed that the highest accuracy was achieved when α_t was in the range of [0.75-0.9] and β_t was in the range of [0.75, 0.8].

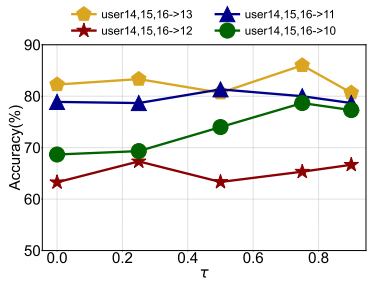


Fig. 16. The sensitivity of hyper-parameter of τ on the Widar3.0 Datasets.

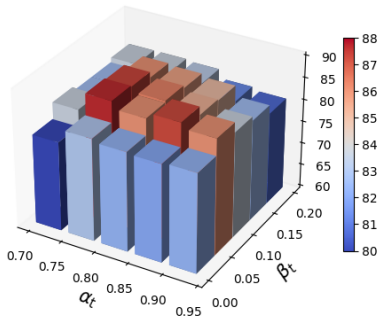


Fig. 17. The sensitivity of the combination of α_s and β_s on the Widar3.0 Datasets. The source domains: user14,15,16. The target domain: user13.

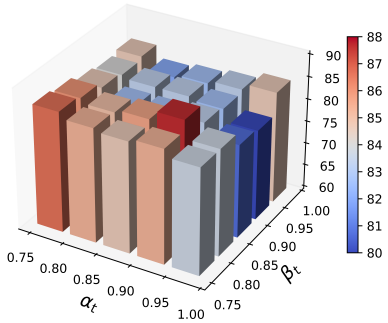


Fig. 18. The sensitivity of the combination of α_t and β_t on the Widar3.0 Datasets.The source domains: user14,15,16. The target domain: user13.