DATTA: Domain-Adversarial Test-Time Adaptation for Cross-Domain WiFi-Based Human Activity Recognition

Julian Strohmayer Rafael Sterzinger Matthias Wödlinger Martin Kampel
Computer Vision Lab, TU Wien
Vienna, Austria
{firstname.lastname}@tuwien.ac.at

Abstract

WiFi-based human activity recognition (HAR) faces significant challenges in cross-domain generalization due to dynamic environmental variations, device heterogeneity, and subtle changes in human behavior. In this paper, we introduce DATTA – Domain-Adversarial Test-Time Adaptation – a novel framework that combines domain-adversarial training (DAT) with test-time adaptation (TTA) and a random weight-resetting mechanism. Unlike previous approaches that apply these techniques in isolation, DATTA is specifically tailored for WiFi-based HAR: it leverages DAT to learn robust, domain-invariant features while TTA continuously refines the model on streaming data. To mitigate catastrophic forgetting during adaptation, we incorporate a weight-resetting mechanism, ensuring sustained performance over prolonged domain shifts. Our extensive experiments on the Widar3.0-G6D dataset demonstrate that DATTA not only outperforms state-of-the-art methods by up to 8.1% in F1-Score but also achieves real-time inference with a lightweight architecture, making it a compelling solution for practical WiFi sensing applications. The PyTorch implementation of DATTA is publicly available at: https://github.com/StrohmayerJ/DATTA.

1. Introduction

WiFi-based sensing has emerged as a promising alternative to optical methods for human activity recognition (HAR) due to its cost-effectiveness, unobtrusiveness, and inherent privacy advantages [7]. These qualities enable efficient, contactless monitoring of human activities in confined indoor environments without the need for per-room sensor deployment, offering a significant economic benefit [30].

However, a critical challenge remains: models trained in one setting often perform poorly when deployed in different environments or on different subjects. This limitation is primarily due to the inherent variability in Channel State Information (CSI), which forms the foundation of modern WiFi-based sensing. CSI, obtained via the Orthogonal Frequency-Division Multiplexing (OFDM) scheme, subdivides a WiFi channel into multiple sub-channels with different carrier frequencies (subcarriers) [13]. This design not only enables fast, parallel data transmission but also provides detailed information on how each subcarrier is affected by the surrounding domain, thereby allowing for the correction of domain-specific noise. By correlating the distinctive patterns of amplitude attenuation and phase shifts in CSI induced by specific human activities, HAR tasks can be effectively performed [22].

Unlike vision-based systems, where domain shifts tend to be more predictable, WiFi signals are extremely sensitive to subtle changes, such as furniture rearrangements, variations in device placements, or minor environmental interferences, which lead to rapid and unpredictable distribution shifts. Here, the term *domain* encompasses factors such as the physical environment, the morphological variability among individuals, differences in sensing hardware, and electromagnetic noise from surrounding devices. Fundamentally, the nature of CSI, designed to capture amplitude and phase perturbations induced by its specific domain, renders models trained on such data inherently domain-specific, resulting in poor cross-domain generalization [1].

Previous methods have primarily focused on domain-invariant feature extraction via domain-adversarial training (DAT) [9]. While DAT helps mitigate overfitting to source-specific characteristics, it falls short when confronted with the rapid and dynamic changes encountered in real-world WiFi domains. In contrast, test-time adaptation (TTA) [34] allows models to adjust on the fly and has been a proven technique in the computer vision domain. However, despite its success, TTA remains largely unexplored in WiFi-based HAR. In our work, we bridge this gap by introducing Domain-Adversarial Test-Time Adaptation (DATTA), a novel framework that integrates DAT with TTA and incorporates random weight resetting [38] to prevent catastrophic forgetting [6]. This combination effectively ensures that the model preserves its learned invariant features while adapting in real time to previously unseen domain characteristics.

Our approach is further strengthened by a tailored augmentation strategy that captures the unique properties of WiFi CSI, along with a lightweight model architecture optimized for rapid adaptation without sacrificing inference speed. Collectively, these techniques address both the fundamental challenge of cross-domain generalization and the practical constraints of real-time operation in WiFi sensing scenarios.

Contributions To address the significant challenge of real-time cross-domain generalization in WiFi-based HAR, we present the following contributions:

(I) We propose Domain-Adversarial Test-Time Adaptation (DATTA), a novel TTA framework that incorporates gradual weight resetting to preserve robust initial features during continuous adaptation.

(II) We adapt the WiFlexFormer architecture for WiFi-based HAR with a CSI-specific augmentation module, enabling efficient adaptation and low-latency inference for real-time deployment.

(III) We validate DATTA through extensive experiments and ablations on public datasets, showing consistent cross-domain gains with minimal computational overhead.

2. Related Work

Cross-Domain WiFi-Based HAR WiFi CSI-based HAR has emerged as a powerful device-free sensing modality, enabling applications in smart homes, healthcare, and security. Yet, cross-domain generalization remains a core obstacle to practical deployment [1]. Recent work addresses this challenge through four key strategies: physics-informed representations, synthetic augmentation, adaptive learning, and large-scale data diversity. Domain-invariant feature extraction reduces sensitivity to domain shifts by exploiting phase differences between antennas [4], constructing body-coordinate velocity profiles in Widar3.0 [43, 47], filtering static interference via AFEE with matching networks [29], applying adversarial training across domains [24, 36], and factorizing denoised, phase-fused CSI into domain-invariant components [2]. Virtual sample generation enhances generalization by synthesizing domain-diverse CSI signals using generative models (GANs [20], VAEs [3]), physically grounded perturbations [28], MixUp [18], and subcarrier dropout [10]. Transfer and few-shot learning approaches minimize labeled data requirements by adapting models to new domains using adversarial training [16, 17], correlation alignment [12], and by leveraging similarity-based models such as DSEN [15], KNN-MMD [44], and CrossFi [45] for one- or zero-shot generalization. Finally, large-scale diversity from multi-link and federated designs, such as WiSDAR [37], WiTraj [39], HandGest [41], and AdaWiFi [46], further improves robustness by exposing models to spatial and contextual variability.

Domain-Adversarial Training In their seminal work on domain-adversarial training, Ganin and Lempitsky [8] introduce the Domain-Adversarial Neural Network (DANN). This architecture utilizes a Gradient Reversal Layer (GRL) to align feature representations between source and target domains by minimizing domain discrepancy in an adversarial manner while preserving discriminative task-related features. An extended version of this work [9] provided further theoretical insights and empirical validation, establishing DANN as a cornerstone in domain adaptation research. Building on DANN, Tzeng et al. [35] improved cross-domain generalization through domain confusion and target-specific classifiers, allowing for more nuanced adaptation to target-specific features. Long et al. [25] further advanced DAT with Conditional Domain-Adversarial Networks (CDAN), which integrate conditional task labels to refine alignment between domains and capture complex dependencies. Peng et al. [27] recently extended DAT with Moment Matching for Multi-Source Domain Adaptation (M^3SDA), which adapts knowledge from multiple labeled source domains to an unlabeled target domain by dynamically aligning higher-order moments of feature distributions. M^3SDA's use of multiple classifiers enables domain-specific adaptation but results in higher inference costs, making it less suitable for real-time applications. In contrast, Jiang et al. [16] presented one of the first successful applications of DAT in WiFi-based HAR. Their approach, combining a single CNN-based feature extractor with an MLP-based domain discriminator and activity recognizer, maintains lower inference times by using a single classifier. This architecture is thus better suited for real-time WiFi-based HAR, which is why we combine it with TTA, as explored in this work, to enable effective adaptation to unseen domains while maintaining computational efficiency.

Test-Time Adaptation TTA methods focus on adapting a pre-trained model during inference without altering its training. Lin et al. [21] introduce a TTA method for RGB video that adjusts test-time statistics to match those from training, promoting consistency across temporally augmented views to enhance coherence in model predictions. Wang et al. [38] address challenges in continual adaptation with posthoc regularization techniques, averaging predictions across model weights and augmentations to reduce error accumulation and restoring model weights to their original state intermittently to prevent catastrophic forgetting. Liu et al. [23] further enhance adaptability by using a self-supervised contrastive learning objective to align source and target features, improving robustness without temporal dependencies. Extending these TTA approaches, Ma et al. [26] propose a so-called improved self-training method that refines pseudo-labels using graph-based correction and stabilizes adaptation with a parameter moving average.

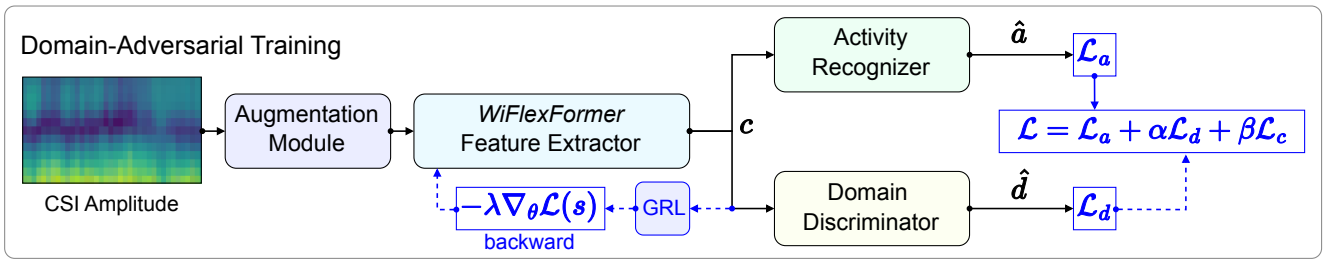

Figure 1. Overview of the domain-adversarial training approach used in DATTA.

3. Domain-Adversarial Test-Time Adaptation

Our proposed framework, DATTA, combines DAT with TTA to enable robust cross-domain generalization in WiFi-based HAR. Through DAT, the model learns domain-invariant features by leveraging data from diverse domains, achieving offline adaptation to varied data characteristics. Our DAT architecture builds on the adversarial structure for domain-invariant feature learning outlined in [16], with a *WiFlexFormer*-based central feature extractor [33] to ensure efficient, real-time HAR. To further enhance cross-domain generalization, we incorporate a specialized augmentation module tailored to the unique properties of WiFi CSI. Despite DAT's effectiveness, residual domain shifts can still occur at test time due to environmental changes that lead to distribution shifts in data. To account for these shifts as well, we adapt the TTA framework from [21] for WiFi-based HAR to align the feature distributions of target and source domains in real-time at test time. Additionally, we leverage random weight resetting [6] to prevent catastrophic forgetting of learned domain invariance during prolonged domain shifts, ensuring sustained model stability.

3.1. Domain-Adversarial Training

Figure 1 illustrates the DAT architecture used in DATTA, consisting of the *Feature Extractor*, *Activity Recognizer*, and *Domain Discriminator*. The feature extractor processes CSI amplitude data to learn domain-invariant features by generating representations that are informative for activity recognition while disregarding domain-specific aspects. During training, the extracted features are fed to the activity recognizer and domain discriminator. The domain discriminator enforces domain invariance by applying an adversarial loss, pushing the feature extractor to produce features that are indistinguishable across domains. After training, the domain discriminator is discarded, leaving a model optimized for cross-domain generalization.

Model Input/Output Our DAT architecture takes CSI amplitude spectrograms $s \in \mathcal{S}$ as input, each associated with an activity label $a \in \mathcal{A}$ and a domain label $d \in \mathcal{D}$, where $\mathcal{A}$ and $\mathcal{D}$ represent the sets of activities and domains, respectively. The output is the predicted activity label $\tilde{a} \in \mathcal{A}$.

Augmentation Module Before feature extraction, raw amplitude spectrograms are processed by the augmentation module, which applies a set of realistic random augmentations to the CSI signal to increase data variability. Among these augmentations are amplitude perturbations, circular rotations along the temporal axis, as well as pixel- and row-wise dropout with mean replacement [31].

Feature Extractor The feature extractor is based on the *WiFlexFormer* [33] architecture which is chosen due to its lightweight design and simplicity. It consists of a convolutional stem followed by Gaussian positional encoding and a transformer encoder with class token and a linear classification layer. Its architecture is designed for both, amplitude and Doppler frequency shift features, however, in our experiments, we will exclusively run in the amplitude mode to minimize the parameter count and inference time. The resulting model is comparatively small with only $\approx$ 40k parameters.

Activity Recognizer The activity recognizer, acting as the classification head in our architecture, consists of two linear layers with a ReLU activation function in between and $|\mathcal{A}|$ output channels. It takes the class token embeddings c as input and outputs the predicted activity probabilities $\hat{a}$. Its loss $\mathcal{L}_a$, the main training objective, is the cross-entropy between the predicted and true activities:

$$\mathcal{L}_a = -\sum_{k=1}^{|\mathcal{A}|} a_k \log(\hat{a}_k). \tag{1}$$

Domain Discriminator The domain discriminator architecture mirrors the activity recognizer with two linear layers with a ReLU activation function in between and $|\mathcal{D}|$ output channels. It takes the class token embeddings c as input and outputs a prediction over domains. The domain loss $\mathcal{L}_d$ is then computed using the cross-entropy between the predicted domain probabilities $\hat{d}$ and the true domain labels d:

$$\mathcal{L}_d = -\sum_{k=1}^{|\mathcal{D}|} d_k \log(\hat{d}_k). \tag{2}$$

Furthermore, to facilitate efficient training without having to freeze model weights alternately, we utilize a GRL.

As shown in Figure 1, during the forward pass, the GRL acts as an identity function, allowing the features to flow unchanged to the domain discriminator. However, during backpropagation, it multiplies the gradients $\nabla_\theta \mathcal{L}(s)$ by $-\lambda$ to reverse them, thus, returning $-\lambda \nabla_\theta \mathcal{L}(s)$ to the feature extractor:

$$\text{Forward Pass:} \qquad \text{GRL}(s) = s, \tag{3}$$

$$\text{Backward Pass:} \qquad \text{GRL}(s) = -\lambda \nabla_\theta \mathcal{L}(s), \tag{4}$$

where λ is a scaling parameter that controls the strength of the adversarial signal. By reversing the gradients, the feature extractor is encouraged to produce features that are indistinguishable across domains, thus learning domain-invariant representations. A more comprehensive description of the GRL can be found in [8].

To perform DAT without overwhelming the feature extractor in the early phase of training, we perform dynamic scaling of λ as follows:

$$\lambda = \Big(\frac{2}{1 + e^{-10p}} - 1 \Big) \gamma, \tag{5}$$

where $p \in [0, 1]$ represents the training progress and γ is a scaling parameter to control the adversarial signal strength. This allows the feature extractor to focus on learning robust features for the primary task of activity recognition in the beginning and, by gradually increasing the strength of the adversarial signal, enables a smooth transition to domain-invariant feature learning.

Domain-Adversarial Loss The loss function in our DAT architecture leverages adversarial training to balance activity recognition and domain-invariant feature learning while penalizing overconfidence in a specific class through a confidence control mechanism. This is achieved by combining task-specific ($\mathcal{L}_a$) and domain-specific ($\mathcal{L}_d$) losses with the axillary loss $\mathcal{L}_c$, representing the *Confidence Control Constraint* (CCC) from [16]. $\mathcal{L}_c$ penalizes predictions that are overly certain by adding a penalty for class probabilities approaching 0 or 1. Here, $\hat{a}_{ik}$ represents the predicted probability for activity class $k \in \mathcal{A}$ of the i-th sample, ensuring that each class prediction is regularized:

$$\mathcal{L}_c = -\sum_{k=1}^{|\mathcal{A}|} \log(\hat{a}_{ik}) + \log(1 - \hat{a}_{ik}). \tag{6}$$

Combined with task- and domain-specific losses, the final domain-adversarial loss function minimized during DAT is given by:

$$\mathcal{L} = \mathcal{L}_a + \alpha \mathcal{L}_d + \beta \mathcal{L}_c, \tag{7}$$

where α and β are weighting parameters, used for controlling the strengths of the adversarial signal and the CCC, respectively.

3.2. Test-Time Adaptation

While DAT is effective in learning domain-invariant features, large domain shifts still lead to a drop in performance during test time. In order to reduce the impact of such domain shifts, we employ TTA, allowing off-the-shelf pretrained models to adapt online to new target domains without requiring additional labeled data.

Building upon the framework proposed by Lin et al. [21], we adapt TTA from RGB video to CSI amplitude spectrograms to further enhance the generalization of DAT during test time by performing feature distribution alignment, i.e., aligning source statistics of the model with online estimates of the target statistics. Additionally, to prevent overfitting to the target distribution during prolonged adaptation, i.e. catastrophic forgetting [6], we implement random weight resetting, following the approach proposed by Wang et al. [38]. Specifically, a subset of model parameters is reverted to their source model values to keep them closer to the domain-invariant feature space.

Feature Map Alignment To address the distribution shift, we align the statistics of feature maps, i.e., matching the means and variances, computed for both the training and test spectrograms. Let $\phi_l(s; \theta)$ represent the feature map of the l-th layer of network ϕ, computed for a spectrogram s with parameters θ. Each feature map is a matrix of dimensions (t_l, f_l), where t_l and f_l correspond to the time steps and frequency channels (subcarrier), respectively.

Computing the mean of the l-th layer features for a dataset $\mathcal{S}$ across the time dimension results in a mean vector of size f_l, which can be expressed as:

$$\mu_l(\mathcal{S}; \theta) = \mathbb{E}_{s \in \mathcal{S}} \mathbb{E}_{t \in [1, t_l]} \Big[\phi_l(x; \theta)[t] \Big], \tag{8}$$

and the variance of the l-th layer features is given by:

$$\sigma_l^2(\mathcal{S}; \theta) = \mathbb{E}_{s \in \mathcal{S}} \mathbb{E}_{t \in [1, t_l]} \Big[(\phi_l(x; \theta)[t] - \mu_l(\mathcal{S}; \theta))^2 \Big]. \tag{9}$$

For the remainder of this work, we denote the mean and variance computed on the training set with $\bar{\mu}_l$ and $\bar{\sigma}_l^2$. When training data is unavailable, these statistics can be estimated from batch norm layers as well, though with a small decrease in performance [21].

At test time, updates are performed iteratively, adjusting the discrepancy between the test statistics of a batch $\mathcal{B}$ of selected layers L with those computed during training:

$$\mathcal{L}_{\text{TTA}} = \sum_{l \in L} ||\mu_l(\mathcal{B}; \theta) - \bar{\mu}_l||_2 + ||\sigma_l^2(\mathcal{B}; \theta) - \bar{\sigma}_l^2||_2. \tag{10}$$

In our experiments, we observe that optimal results are obtained by selecting L to contain only the first out of four transformer encoder layers that compose the *WiFlexFormer* encoder.

Given the low inference time of the *WiFlexFormer*, we perform TTA for the most realistic application scenario: online, on data received in a stream. Hence, we chose $|\mathcal{B}| = 1$

and continuously evaluate target statistics using exponential moving averages instead of repeatedly computing statistics for the constantly growing test set. In other words, given the spectrogram s_i, received in iteration i, we update the mean and variance estimates as follows:

$$\hat{\mu}_l^{(i)} = \alpha \cdot \mu_l(s_i;\theta) + (1-\alpha)\cdot \hat{\mu}_l^{(i-1)}, \tag{11}$$

$$\hat{\sigma}_l^{2(i)} = \alpha \cdot \sigma_l^2(s_i;\theta) + (1-\alpha)\cdot \hat{\sigma}_l^{2(i-1)}, \tag{12}$$

where $1-\alpha$ denotes the momentum. As a starting point, we select the source statistics, i.e., $\mu_l^{(0)} = \bar{\mu}_l$ and $\sigma_l^{2(0)} = \bar{\sigma}_l^2$. Without modifications to Eq. 10 and with no extensive recomputation necessary, we compute $\mathcal{L}_{\text{TTA}}$ using these estimates instead.

Weight Resetting In order to avoid catastrophic forgetting, in each iteration, we reset a subset of the current weights to their original values from $\bar{\theta}$, the source models's parameters. Specifically, consider the weights of layer l in iteration i, denoted as $\theta_l^{(i)}$. To randomly reset these, we define a Boolean mask m_l with $\dim(\theta_l^{(i)}) = \dim(m_l)$, where each element of the mask is sampled from a Bernoulli distribution with reset rate p. The updated parameter vector $\theta_l^{(i)}$ is then:

$$\theta_l^{(i)} = m_l \odot \bar{\theta}_l + (1-m_l)\odot \theta_l^{(i)}, \tag{13}$$

where $\odot$ denotes element-wise multiplication.

4. Evaluation

We evaluate DATTA's effectiveness through a detailed ablation study of its components, including the augmentation module, loss function elements, and discriminator input in DAT, as well as random weight resetting in TTA to assess its impact on model stability. Our experiments use a modified version of a publicly available dataset adapted for DAT and TTA. We also analyze inference time to assess the computational impact of TTA and weight resetting on real-time performance, addressing practical deployment feasibility.

4.1. Data

Widar3.0-G6D Dataset The Widar3.0-G6D dataset is the primary benchmark used in our evaluation of cross-domain generalization. It is derived from the Widar3.0 dataset [42], which contains CSI recordings of 22 hand gestures performed by 16 participants across three indoor environments. Since not all gestures are uniformly present in all settings, we adopt the Widar3.0-G6 subset [14], which includes 6 gestures (*push and pull*, *sweep*, *clap*, *slide*, *draw circle*, *draw zigzag*) consistently recorded across all environments by all users, yielding ≈ 68k gesture samples. The recording setup consists of a single transmitter and six receivers, each equipped with an Intel WiFi Link 5300 NIC operating at 5 GHz, collecting CSI from 3 antennas × 30 subcarriers at 1,000 Hz using the Linux CSI Tool [11].

Subset	Activities	Envs.	Persons	Domains	Samples
Train	6	2	7	7	19,586
Val	6	2	7	7	4,896
Val_{TTA}	6	1	9	9	3,417
Test	6	1	9	9	30,749
Total	6	3	16	16	58,648

Table 1. Overview of the distribution of activities, environments, participants, domains, and activity samples across the Widar3.0-G6D subsets used for training, validation, and testing.

To construct Widar3.0-G6D, we extract CSI from 30 subcarriers of the first antenna at each receiver, apply temporal sub-sampling to 100 Hz, and retain samples between 120 and 220 packets (1.2–2.2 s), zero-padding shorter sequences. Amplitude features are normalized via min-max scaling. The final dataset comprises ≈ 58k samples, split into disjoint training and test subsets based on unique room-participant combinations: training uses 7 domains (2 environments, 7 users), and testing uses 9 domains (1 environment, 9 users), as shown in Table 1. This setup provides a controlled yet diverse domain structure for evaluating generalization performance.

3DO Dataset The 3DO dataset [32] is a through-wall WiFi HAR dataset designed to isolate environmental domain shifts while eliminating inter-subject variability. It includes three macroscopic activities (*walking*, *sitting*, *lying*) performed by a single participant over three days in a 6 m × 5 m office, with a single transmitter and receiver (*ESP32-S3-DevKitC-1U*, 2.4 GHz, 100 Hz) placed in adjacent rooms 7.2 m apart. Day 1 provides in-domain data, day 2 introduces dynamic and temporal variation, and day 3 adds static environmental changes. Each day includes five 5-minute sequences per activity, totaling over 1.2 million labeled WiFi packets. To evaluate cross-domain generalization, we merge data from days 1 and 2 and apply an 8:2 training/validation split. Data from day 3 is reserved for testing, creating a controlled domain shift focused solely on environmental factors. Samples have a shape of 52 × 351 (≈ 3.5 seconds).

MM-Fi Dataset The MM-Fi dataset [40] is a multi-modal HAR dataset featuring synchronized recordings from WiFi CSI, LiDAR, mmWave radar, and RGB-D video. It includes 27 activities (14 daily, 13 rehabilitation) performed by 40 subjects across four indoor environments. WiFi data is collected using TP-Link N750 routers with the Atheros CSI Tool from a 1×3 MIMO link at 5 GHz, 40 MHz bandwidth, with the transmitter placed 0.75 m from the subject and the 3-antenna receiver 3.0 m away. CSI is recorded at 1000 Hz and downsampled to 100 Hz, yielding ≈ 320k WiFi packets. To evaluate cross-domain generalization, we construct disjoint subsets based on environment-participant combinations: training/validation data (8:2 split) from three environ-

Model	A	R	ACC	F1-Score
W (baseline)	-		38.69 ±3.17	40.62 ±2.93
	✓		47.75 ±4.31	49.32 ±4.45
W_{TTA}	-	-	42.76 ±4.21	44.89 ±3.93
	✓	-	51.70 ±4.40	53.20 ±4.26
W_{DAT}	-		39.54 ±2.80	42.02 ±2.62
	✓		64.53 ±1.87	65.66 ±1.85
W_{DATTA}	-	-	45.26 ±5.83	48.23 ±5.11
	✓	-	65.90 ±1.53	67.29 ±1.43
	✓	✓	66.92 ±1.54	68.13 ±1.45

Table 2. Cross-domain HAR performance on the Widar3.0-G6D dataset averaged over three runs, comparing models trained with conventional training (W), test-time adaptation (W_{TTA}), domain-adversarial training (W_{DAT}), and domain-adversarial test-time adaptation (W_{DATTA}). Columns A and R indicate the use of data augmentation and random weight resetting during TTA.

ments and 30 participants, and test data from the remaining environment and 10 participants. We select six activities (A8, A9, A10, A11, A16, A25) and extract CSI from the first antenna pair, resulting in samples of shape 114×297 (≈ 3 seconds).

4.2. Model Training

We evaluate four model configurations to assess the impact of DATTA and its components: the baseline *WiFlexFormer* (W), its TTA variant (W_{TTA}), the DAT model (W_{DAT}), and the final DATTA model (W_{DATTA}). All model backbones are trained on *Train*, validated on *Val*, with TTA hyperparameters tuned on Val_{TTA}, and tested on the shuffled dataset *Test*, if not stated otherwise (cf. Table 1). Note that shuffling *Test* poses an extreme case of high-frequency domain shifts. A Python script that creates these exact subsets from Widar3.0-G6 is provided for reproducibility.

The baseline W uses the vanilla *WiFlexFormer* architecture without DAT or TTA. W_{DAT} incorporates DAT with loss weights $\alpha = 0.3$, $\beta = 0.2$, and GRL-scaling $\gamma = 8$. Adding TTA to W and W_{DAT} produces W_{TTA} and W_{DATTA}, allowing the model's first layer to adapt during inference. The complete DATTA framework, W_{DATTA}, is tuned using Bayesian and grid search. For all models using TTA, random weight resetting with $p = 1 \times 10^{-4}$ is evaluated to enhance stability. We further assess our augmentation module's effectiveness by training each model with and without augmentations. Detailed hyperparameters for all model configurations are provided in the supplementary material.

4.3. Results

Augmentation Module We begin our evaluation with the augmentation module, the first component in our DAT pipeline. Inspecting Table 2, which depicts all cross-domain activity recognition results, reveals the critical role of our data augmentation approach for handling cross-domain variations. Comparing the baseline model W with and without augmentation, we observe a substantial improvement in F1-Score, from 40.62% to 49.32%.

Its impact is even more pronounced in W_{DAT}, the DAT model: without augmentation, W_{DAT} achieves an F1-Score of only 42.02%, reflecting poor generalization and significant overfitting to domain-specific features. Hence, augmentation is crucial for DAT as without it, the model overfits early and fails to learn domain-invariant features at later stages of training when the adversarial signal strength is increased. However, with augmentation, W_{DAT} attains an F1-Score of 65.66%, showing that sufficient data variability is essential to facilitating domain-invariant feature learning.

Model	CCC	ACC	F1-Score
W_{DAT}	-	62.81 ±1.09	63.81 ±1.09
W_{DAT}	✓	64.53 ±1.87	65.66 ±1.85
W_{DATTA}	-	65.62 ±1.08	66.78 ±1.08
W_{DATTA}	✓	66.92 ±1.54	68.13 ±1.45

Table 3. Ablation study on the CCC, weighted with $\beta = 0.2$.

Confidence Control Constraint Next, we consider the impact of the CCC which is designed to prevent overconfidence in specific class predictions to encourage a more balanced feature representation across classes. By discouraging extreme confidence in particular classes, CCC helps stabilize training, making the model more adaptable to unseen domains. By means of a hyperparameter search, we identified the optimal weight $\beta = 0.2$ for CCC and evaluate its effectiveness, we conducted an ablation study comparing models trained with and without CCC (i.e., $\beta = 0$). Observing the results, shown in Table 3, shows that including a CCC improves performance in both the W_{DAT} and W_{DATTA} models. For W_{DAT}, enabling CCC leads to an increase in F1-Score from 63.81% to 65.66%, and for W_{DATTA}, from 65.62% to 68.13%, reinforcing that CCC plays a valuable role in enhancing cross-domain generalization.

Random Weight Resetting Following this, we evaluate the impact of random weight resetting during TTA which has been introduced to prevent catastrophic forgetting. Our results, presented in Table 2, indicate that random weight resetting promotes cross-domain generalization, however only when the model is domain-invariant to some degree, as achieved through DAT with data augmentation.

For W_{DATTA} with augmentation, enabling random weight resetting boosts the F1-Score to 68.13%, up from 67.29% without resetting, achieving the highest performance within our model landscape. However, in other cases, it has limited or even negative effects. For instance, applying weight resetting to W_{DATTA} without augmentation or to W_{TTA} leads to reduced performance. We hypothesize that this is due to the weights of the base model W and W_{DAT} without augmen-

tation not being sufficiently domain-invariant yet, causing weight resets to revert beneficial adaptations during TTA. In contrast, W_{DATTA}, trained with augmentation and thus exhibiting stronger domain invariance, allows weight resetting to maintain proximity to this invariant space during TTA. Consequently, it prevents drifting due to over-adaptation to specific test domains, avoiding catastrophic forgetting and stabilizing performance across diverse domains.

Cross-Domain Adaptation Moving forward, Figure 2 illustrates the performance of three models: DATTA with weight resetting ($W_{\text{DATTA+R}}$), DATTA without weight resetting (W_{DATTA}), and DAT without TTA as the baseline (W_{DAT}), across three domain sequences to evaluate adaptability and resilience to domain shifts.

In the first experiment (top plot), domains are processed sequentially (D0 to D8) to simulate realistic domain shifts. TTA consistently improves performance through incremental adaptation, except for domain D8, which experiences a notable decline. PCA analysis reveals that D8 has low intra-class cohesion and poor separability, indicating high label noise and ambiguity. Severe class imbalance further supports labeling issues in D8, as class 1 comprises only 0.3% of samples (8 out of ∼2,700), unlike domains D0–D7 with uniform class distributions (see supplementary material Section 2). These factors explain the accuracy drop in D8 due to systematic misclassification. However, TTA effectively adapts to other domains, confirming its capability to overcome participant-induced domain gaps.

In the second experiment (middle plot), we process the domains in reversed order (D8 to D0) to test the ability to recover from domains with stark differences in underlying data statistics. After finishing adaptation to the ill-posed domain, D8, TTA quickly recovers and exceeds baseline performance, demonstrating its resilience in recovering from challenging domains.

The third experiment (bottom plot) assesses resilience to catastrophic forgetting by alternating between domains D0 and D2, simulating antagonistic domain shifts. As shown previously, D2 (like D8) negatively affects performance, unlike other domains. This configuration thus represents a challenging setting. Here, $W_{\text{DATTA+R}}$ maintains stable performance across shifts, showcasing robustness in retaining learned features. Conversely, W_{DATTA} shows continual performance degradation, suggesting gradual loss of domain invariance and highlighting the effectiveness of weight resetting to prevent catastrophic forgetting.

Overall, both $W_{\text{DATTA+R}}$ and W_{DATTA} significantly outperform the baseline W_{DAT}, confirming TTA's effectiveness in enhancing cross-domain generalization. Additionally, $W_{\text{DATTA+R}}$ outperforms W_{DATTA} across all experiments, validating the role of random weight resetting in preventing performance degradation under repeated, especially prolonged domain shifts.

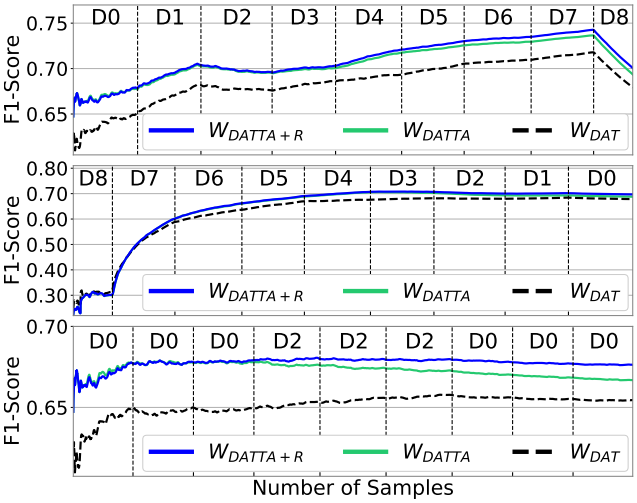

Figure 2. TTA performance across continuous test domain sequences. From top to bottom: (1) ascending domain order (D0 to D8), (2) descending domain order (D8 to D0), and (3) alternating domain order with prolonged domains D0 and D2. Depicted are F1-Scores computed with a rolling window of 100 samples for DATTA models with weight resetting ($W_{\text{DATTA+R}}$, blue), without weight resetting (W_{DATTA}, green), and the baseline DAT model without TTA (W_{DAT}, black).

SotA Comparison To compare DATTA with existing SotA methods, we evaluate it against the default DAT model as proposed in [16] and the video-based TTA variant in [21]. To the best of our knowledge, we are the first to port this video-based TTA variant to the WiFi domain, presenting challenges for direct comparison. To align as closely as possible with the original work, we retain all parameters, tuning only the layers for statistic alignment (choosing the first layer instead of the last two) and the learning rate (1×10^{-6}).

Shown in Table 4 are the results of our SotA evaluation; note that all variants have been trained using the suggested data augmentation module. When combining ViTTA [21] with W or W_{DAT}, the F1-Score stays almost the same ($+0.4\%/-0.1\%$), indicating that employing the method as-is has no benefit. Consequently, evaluating it against our TTA variants, namely W_{TTA} and W_{DATTA} without weight resetting, performance is worse (49.54% versus 52.2% and 67.29%), suggesting that adopting ViTTA to the WiFi domain is not straightforward, requiring changes in the loss function, target initialization, and extensive parameter tuning. Similarly, when combining DAT [16] with W or W_{TTA}, performance is significantly improved over W, though remains below our DAT variants (63.16% vs. 65.66% and 67.29%), i.e., W_{DAT} and W_{DATTA}. Here, the key difference is the input to the discriminator: class token embeddings c concatenated with ([16]) and without (ours) activity logits $\hat{a}$. We hypothesize that this is due to $\hat{a}$ inherently including domain information, namely priors over activities, e.g., lying being more likely performed in the bedroom than in the office. Hence, by including $\hat{a}$, the activity recognizer

Model		ACC	F1-Score
W	+ViTTA [21]	47.76 ±4.33	49.54 ±4.43
W	+DAT [16]	61.87 ±1.39	63.16 ±1.55
W_{TTA}	+DAT [16]	64.48 ±1.58	65.87 ±1.41
W_{DAT}	+ViTTA [21]	64.37 ±2.05	65.60 ±1.94
W	+both	61.66 ±0.81	63.04 ±0.92
W_{DATTA}		66.92 ±1.54	68.13 ±1.45

Table 4. SotA comparison against DAT [16] and ViTTA [21]. (W and W_{TTA}) + DAT use the discriminator input $c \oplus \hat{a}$, (W and W_{TTA}) + ViTTA, employ the ℓ_1-loss, and initialize targets with zero, with slightly tuned hyperparameters. All models are trained with augmentation, and W_{DATTA} additionally uses weight resetting.

is penalized for learning these priors, making predictions less accurate. Finally, we compare combining both SotA methods naively (W+both) against our method W_{DATTA}: as before, the impact of ViTTA is negligible, resulting in an improvement over the SotA by 8.1%.

To evaluate the generalizability of DATTA beyond Widar3.0-G6D, we assess its cross-domain performance on two additional datasets: 3DO and MM-Fi. We compare it against SotA non-adaptive methods developed for WiFi/RF signal processing, namely THAT [19] and RF-Net [5]. As shown in Table 5, DATTA consistently outperforms all baselines across both datasets. On 3DO, our base model W performs comparably to THAT and RF-Net, while W_{DATTA} achieves an improvement of +4.65 F1-Score points, demonstrating the effectiveness of DATTA. On MM-Fi, the advantage is even more pronounced: while non-adaptive models fail to generalize and perform close to random guessing ($\frac{1}{6}$), W_{DATTA} surpasses the next best model W by 12.68 F1-Score points. These results confirm that DATTA transfers effectively across WiFi standards (2.4 GHz vs. 5 GHz), deployment scenarios (line-of-sight vs. through-wall), activity scale (micro- vs. macroscopic), and CSI characteristics (subcarrier configuration, sampling rate, window size, etc.).

Inference Time Table 6 compares the inference times for the baseline model W, W_{DAT}, W_{TTA}, W_{DATTA}, and non-adaptive methods THAT and RF-Net on both an Nvidia RTX 2070 GPU and a Jetson Orin Nano single-board computer. On the RTX 2070, W achieves the fastest inference time at 1.98 ms per sample, with a minor increase to 2.05 ms for W_{DAT} due to the DAT component. Models incorporating TTA show a notable rise in inference time due to additional adaptation steps, with W_{TTA} reaching 9.86 ms and W_{DATTA} reaching 10.05 ms, a $\approx$ 5x increase over the baseline. With random weight resetting, inference time increases further to approximately 19.52 ms for W_{TTA} and 20.21 ms for W_{DATTA}, roughly a 10x increase. Notably, while models like THAT and RF-Net have significantly higher parameter counts (1.7M–3.2M) and inference times, our base model W, having only 40.8k parameters, achieves comparable performance on 3DO and MM-Fi.

	3DO		MM-Fi	
Model	ACC	F1-Score	ACC	F1-Score
THAT [19]	70.62 ±1.43	70.60 ±1.42	19.79 ±2.64	20.63 ±3.80
RF-Net [5]	73.66 ±0.71	73.37 ±0.68	19.87 ±4.80	20.16 ±7.43
W	72.83 ±0.13	72.81 ±0.15	22.72 ±3.35	20.34 ±5.76
W_{DATTA}	77.22 ±3.04	78.02 ±2.93	31.67 ±1.36	33.02 ±2.52

Table 5. Cross-domain adaptation on 3DO and MM-Fi.

			Inference Time [ms]	
Model	**Params.**	**R**	RTX 2070	JON
THAT [19]	3.19 M		8.64 ±0.60	39.53 ±0.58
RF-Net [5]	1.71 M		2.83 ±0.64	14.61 ±0.99
W	40.80 k		1.98 ±0.21	9.24 ±0.25
W_{DAT}	41.97 k		2.05 ±0.23	10.30 ±0.18
W_{TTA}	40.80 k	-	9.86 ±1.42	56.73 ±1.29
	40.80 k	✓	19.52 ±2.27	111.39 ±3.93
W_{DATTA}	41.97 k	-	10.05 ±1.40	57.63 ±1.31
	41.97 k	✓	20.21 ±2.05	115.39 ±3.76

Table 6. Inference time comparison for the base model W, TTA model W_{TTA}, DAT model W_{DAT}, DATTA model W_{DATTA}, and the non-adaptive methods THAT and RF-Net. Column R indicates the use of random weight resetting. Mean inference time is reported over 1,000 iterations (after 100 warm-up iterations) with batch size 1 on an Nvidia RTX 2070 GPU and a Jetson Orin Nano (JON) single-board computer.

Switching from the RTX 2070 to the Jetson Orin Nano incurs an additional $\approx$ 5x increase in inference time across models, with W_{DATTA} achieving around 115.39 ms per sample with weight resetting. However, this latency remains sufficient for HAR, achieving approximately 9 frames per second. Although weight resetting adds to inference time, it significantly improves model stability by preventing catastrophic forgetting, which is crucial for sustained performance across domain shifts. Notably, our TTA code is not optimized for speed, and further optimizations could reduce adaptation time.

5. Conclusion

This work introduced DATTA, a framework for improving cross-domain generalization in WiFi-based HAR by combining domain-adversarial training with test-time adaptation. To preserve robustness during adaptation, DATTA incorporates random weight resetting, which mitigates catastrophic forgetting. It builds on a lightweight Transformer architecture, enabling real-time performance on edge devices. Extensive ablation studies validate the contribution of each component, and demonstrate consistent gains in generalization across a range of public datasets. To support future research, we release a PyTorch implementation of DATTA, along with tools for dataset preparation and evaluation.

Acknowledgments This research was supported by the iMars project MSCA (grant agreement no. 101182996).

References

[1] Chen Chen, Gang Zhou, and Youfang Lin. Cross-domain wifi sensing with channel state information: A survey. *ACM Computing Surveys*, 55(11):1–37, 2023. 1, 2

[2] Xingcan Chen. Cross-domain human activity recognition using reconstructed wi-fi signal. *Physical Communication*, 71: 102651, 2025. 2

[3] Xi Chen, Hang Li, Chenyi Zhou, Xue Liu, Di Wu, and Gregory Dudek. Fidora: Robust wifi-based indoor localization via unsupervised domain adaptation. *IEEE Internet of Things Journal*, 9(12):9872–9888, 2022. 2

[4] Enjie Ding, Xiansheng Li, Tong Zhao, Lei Zhang, Yanjun Hu, et al. A robust passive intrusion detection system with commodity wifi devices. *Journal of Sensors*, 2018, 2018. 2

[5] Shuya Ding, Zhe Chen, Tianyue Zheng, and Jun Luo. Rf-net: A unified meta-learning framework for rf-enabled one-shot human activity recognition. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 517–530, 2020. 8

[6] Sayna Ebrahimi, Franziska Meier, Roberto Calandra, Trevor Darrell, and Marcus Rohrbach. Adversarial Continual Learning. In *Computer Vision – ECCV 2020*, pages 386–402. Springer International Publishing, Cham, 2020. Series Title: Lecture Notes in Computer Science. 1, 3, 4

[7] Biying Fu, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Sensing technology for human activity recognition: A comprehensive survey. *IEEE Access*, PP:1–1, 2020. 1

[8] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015. 2, 4

[9] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016. 1, 2

[10] Kaixuan Gao, Huiqiang Wang, Hongwu Lv, and Wenxue Liu. Toward 5g nr high-precision indoor positioning via channel frequency response: A new paradigm and dataset generation method. *IEEE Journal on Selected Areas in Communications*, 40(7):2233–2247, 2022. 2

[11] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Tool release: Gathering 802.11n traces with channel state information. *Computer Communication Review*, 41: 53, 2011. 5

[12] Zhanjun Hao, Juan Niu, Xiaochao Dang, and Danyang Feng. Wi-cal: A cross-scene human motion recognition method based on domain adaptation in a wi-fi environment. *Electronics*, 11(16):2607, 2022. 2

[13] Steven M. Hernandez and Eyuphan Bulut. Wifi sensing on the edge: Signal processing techniques and challenges for real-world systems. *IEEE Communications Surveys & Tutorials*, 25(1):46–76, 2023. 1

[14] Weiying Hou and Chenshu Wu. Rfboost: Understanding and boosting deep wifi sensing via physical data augmentation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8:1 – 26, 2024. 5

[15] Pengli Hu, Chengpei Tang, Kang Yin, and Xie Zhang. Wigr: a practical wi-fi-based gesture recognition system with a lightweight few-shot network. *Applied Sciences*, 11(8):3329, 2021. 2

[16] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. Towards environment independent device free human activity recognition. In *Proceedings of the 24th annual international conference on mobile computing and networking*, pages 289–304, 2018. 2, 3, 4, 7, 8

[17] Hua Kang, Qian Zhang, and Qianyi Huang. Context-aware wireless-based cross-domain gesture recognition. *IEEE Internet of Things Journal*, 8(17):13503–13515, 2021. 2

[18] Hoonyong Lee, Changbum R. Ahn, and Nakjung Choi. Toward single occupant activity recognition for long-term periods via channel state information. *IEEE Internet of Things Journal*, pages 1–1, 2023. 2

[19] Bing Li, Wei Cui, Wei Wang, Le Zhang, Zhenghua Chen, and Min Wu. Two-stream convolution augmented transformer for human activity recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 286–293, 2021. 8

[20] Xinyi Li, Liqiong Chang, Fangfang Song, Ju Wang, Xiaojiang Chen, Zhanyong Tang, and Zheng Wang. Crossgr: Accurate and low-cost cross-target gesture recognition using wi-fi. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(1), 2021. 2

[21] Wei Lin, Muhammad Jehanzeb Mirza, Mateusz Kozinski, Horst Possegger, Hilde Kuehne, and Horst Bischof. Video Test-Time Adaptation for Action Recognition. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22952–22961, Vancouver, BC, Canada, 2023. IEEE. 2, 3, 4, 7, 8

[22] Jiao Liu, Guanlong Teng, and Feng Hong. Human activity sensing with wireless signals: A survey. *Sensors*, 20(4), 2020. 1

[23] Yuejiang Liu, Parth Kothari, Bastien Van Delft, Baptiste Bellot-Gurlet, Taylor Mordan, and Alexandre Alahi. Ttt++: When does self-supervised test-time training fail or thrive? *Advances in Neural Information Processing Systems*, 34: 21808–21820, 2021. 2

[24] Yan Liu, Anlan Yu, Leye Wang, Bin Guo, Yang Li, Enze Yi, and Daqing Zhang. Unifi: A unified framework for generalizable gesture recognition with wi-fi signals using consistency-guided multi-view networks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 7(4), 2024. 2

[25] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31, 2018. 2

[26] Jing Ma. Improved self-training for test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23701–23710, 2024. 2

[27] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source

domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1406–1415, 2019. 2

[28] Omer Gokalp Serbetci, Ju-Hyung Lee, Daoud Burghal, and Andreas F. Molisch. Simple and effective augmentation methods for csi based indoor localization, 2023. 2

[29] Zhenguo Shi, Qingqing Cheng, J Andrew Zhang, and Richard Yi Da Xu. Environment-robust wifi-based human activity recognition using enhanced csi and deep learning. *IEEE Internet of Things Journal*, 9(24):24643–24654, 2022. 2

[30] Julian Strohmayer and Martin Kampel. Wifi csi-based long-range through-wall human activity recognition with the esp32. In *Computer Vision Systems*, pages 41–50, Cham, 2023. Springer Nature Switzerland. 1

[31] Julian Strohmayer and Martin Kampel. Wifi CSI-based long-range person localization using directional antennas. In *The Second Tiny Papers Track at ICLR 2024*, 2024. 3

[32] Julian Strohmayer and Martin Kampel. On the generalization of wifi-based person-centric sensing in through-wall scenarios. In *Pattern Recognition*, pages 194–211, Cham, 2025. Springer Nature Switzerland. 5

[33] Julian Strohmayer, Matthias Wödlinger, and Martin Kampel. Wiflexformer: Efficient wifi-based person-centric sensing, 2024. 3

[34] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pages 9229–9248. PMLR, 2020. 1

[35] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017. 2

[36] Dazhuo Wang, Jianfei Yang, Wei Cui, Lihua Xie, and Sumei Sun. Airfi: Empowering wifi-based passive human gesture recognition to unseen environment via domain generalization. *IEEE Transactions on Mobile Computing*, 23(2):1156–1168, 2024. 2

[37] Fangxin Wang, Wei Gong, and Jiangchuan Liu. On spatial diversity in wifi-based human activity recognition: A deep learning-based approach. *IEEE Internet of Things Journal*, 6 (2):2035–2047, 2018. 2

[38] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual Test-Time Domain Adaptation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7191–7201, New Orleans, LA, USA, 2022. IEEE. 1, 2, 4

[39] Dan Wu, Youwei Zeng, Ruiyang Gao, Shenjie Li, Yang Li, Rahul C Shah, Hong Lu, and Daqing Zhang. Witraj: Robust indoor motion tracking with wifi signals. *IEEE Transactions on Mobile Computing*, 22(5):3062–3078, 2021. 2

[40] Jianfei Yang, He Huang, Yunjiao Zhou, Xinyan Chen, Yuecong Xu, Shenghai Yuan, Han Zou, Chris Xiaoxuan Lu, and Lihua Xie. Mm-fi: Multi-modal non-intrusive 4d human dataset for versatile wireless sensing. *Advances in Neural Information Processing Systems*, 36:18756–18768, 2023. 5

[41] Jie Zhang, Yang Li, Haoyi Xiong, Dejing Dou, Chunyan Miao, and Daqing Zhang. Handgest: Hierarchical sensing for robust-in-the-air handwriting recognition with commodity wifi devices. *IEEE Internet of Things Journal*, 9(19): 19529–19544, 2022. 2

[42] Yi Zhang, Yue Zheng, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. Widar3.0: Zero-effort cross-domain gesture recognition with wi-fi. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 8671–8688, 2022. 5

[43] Yi Zhang, Yue Zheng, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. Widar3.0: Zero-effort cross-domain gesture recognition with wi-fi. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 8671–8688, 2022. 2

[44] Zijian Zhao, Zhijie Cai, Tingwei Chen, Xiaoyang Li, Hang Li, and Guangxu Zhu. Knn-mmd: Cross domain wi-fi sensing based on local distribution alignment. *arXiv preprint arXiv:2412.04783*, 2024. 2

[45] Zijian Zhao, Tingwei Chen, Zhijie Cai, Xiaoyang Li, Hang Li, Qimei Chen, and Guangxu Zhu. Crossfi: A cross domain wi-fi sensing framework based on siamese network. *arXiv preprint arXiv:2408.10919*, 2024. 2

[46] Naiyu Zheng, Yuanchun Li, Shiqi Jiang, Yuanzhe Li, Rongchun Yao, Chuchu Dong, Ting Chen, Yubo Yang, Zhimeng Yin, and Yunxin Liu. Adawifi, collaborative wifi sensing for cross-environment adaptation. *IEEE Transactions on Mobile Computing*, 24(2):845–858, 2025. 2

[47] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. Zero-effort cross-domain gesture recognition with wi-fi. In *Proceedings of the 17th annual international conference on mobile systems, applications, and services*, pages 313–325, 2019. 2