

RFF-TTA: Physical Information-Aware Prototype for Temporally Varying RF Fingerprinting Online Test-Time-Adaptation

Anonymous submission

Abstract

In recent years, deep learning (DL)-based RF fingerprint (RFF) recognition has become a promising wireless device verification technique in the Internet of Things (IoT). However, temporal variations in device load status effects as well as channel effects can lead to inconsistent RF fingerprint distributions during the training and testing phases, which causes performance degradation of DL models. To address this problem, we propose the first test-time-adaptation (TTA) approach to improve the domain generalization ability of RFF recognition models. We first analyze the causes of time-varying RFF distribution shifts, such as carrier frequency offset (CFO), and develop a physical impairment-based data augmentation strategy. Based on this, we further propose a physically information-aware prototype to guide the model for TTA. Our method requires no model retraining or labeled test samples, and is a lightweight, nonparametric solution. Finally, our approach is extensively evaluated using mobile phones with the IEEE 802.11 orthogonal frequency division multiplexing (OFDM) system, which demonstrates that our scheme can effectively improve RFF average recognition performance by about 7.8%.

Introduction

The proliferation of Internet of Things (IoT) devices has significantly advanced the development of smart homes and smart city applications. Nevertheless, the inherent openness of wireless communication channels renders these connected devices particularly vulnerable to security threats, including unauthorized access. Recent progress in RF fingerprinting (RFF) has established it as a promising authentication mechanism for IoT systems (Huang et al. 2017; Polak and Goeckel 2015), utilizing intrinsic hardware imperfections like IQ imbalance (Huang et al. 2017), loop filter variations (Brik et al. 2008), and clock jitter (Jana and Kasera 2008) as unique, tamper-resistant identifiers.

In recent years, deep learning (DL)-based RFF recognition techniques have been widely investigated, which can achieve impressive classification accuracy (Riyaz et al. 2018; Shen et al. 2022). However, most DL approaches assume ideal conditions with stable, independent, and identically distributed (i.i.d.) for RFF. Real-world deployments face significant domain shifts due to environmental changes and device status variations as shown in Fig. 1, leading to

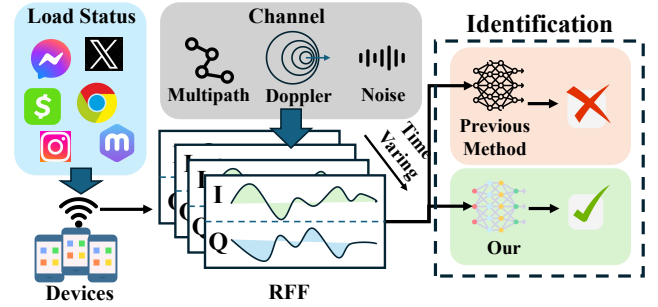


Figure 1: Our work addresses the more realistic scenario where complex device loads (e.g., on a mobile phone) induce DSE, distorting the RFF and degrading accuracy in the test phase.

degraded generalization performance. This presents two key challenges:

Channel Effect. Wireless communication environments exhibit dynamic channel conditions and noise levels that vary with time, location, and acquisition settings. These variations cause significant shifts in signal characteristics (amplitude, frequency, and phase), ultimately degrading the classification performance of deep learning models.

Device Status Effect (DSE). Wireless devices experience varying operational states (e.g., various application loads lead to temperature differences) that significantly affect their RF fingerprints (Li et al. 2024). As shown in Fig. 4c, even fundamental RFF characteristics like carrier frequency offset (CFO) and carrier phase offset (CPO) exhibit notable distribution shifts across different mobile applications for the same device.

Existing studies have primarily focused on mitigating channel effects, proposing methods such as extracting channel-independent features (Liu et al. 2019; Restuccia et al. 2019), developing channel-robust features through transmitter operation (Rajendran and Sun 2022; Rajendran et al. 2020), and applying advanced signal processing algorithms (Xing et al. 2022; Shen et al. 2022). However, these algorithms do not take into account the phenomenon of RFF offset in the device itself due to complex DSE. To cope with the DSE, the researchers explored the effect of temperature on the CFO fingerprints in mobile phone (Li et al. 2024).

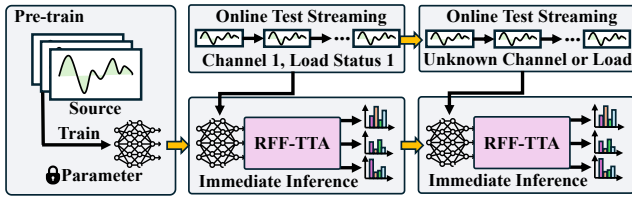


Figure 2: We propose *RFF-TTA* to solve this DG problem. *RFF-TTA* can utilize batch samples from an online test stream to guide model adaptation and direct inference.

Then, the research attempts to address this problem by modeling recognition at different phone temperatures (Gu et al. 2024). However, the prior temperature of the phone is unavailable in many cases. Furthermore, the combined effects of channel variations and DSE lead to highly complex and unpredictable RFF distribution patterns, significantly compounding the identification challenges.

Recent studies have attempted to address the RFF distribution shift problem through domain adaptation (DA) (Zhang et al. 2024; Cai et al. 2025) and domain generalization (DG) (Tang et al. 2024) techniques. However, DG struggles with unseen device states, while DA requires test data for training - both are impractical for online scenarios. In the broader machine learning field, test-time adaptation (TTA) has emerged as an effective paradigm to address distribution shifts by dynamically adjusting models using unlabeled test data. This approach is particularly appealing for RFF recognition as it enables real-time adaptation without costly retraining. However, existing TTA strategies (Jiang et al. 2024; Gong et al. 2025), primarily developed for computer vision tasks, demonstrate limited effectiveness when directly applied to RF fingerprint data due to fundamental differences in data characteristics. For example, as shown in Fig. 4a, the latest decoupled prototype learning (DPL) (Wang et al. 2025) and non-parametric classifier (NPC) (Zhang et al. 2023) methods are unstable and have even experienced a decline compared to the DG method.

This paper presents *RFF-TTA*, a novel TTA method that uniquely exploits inherent DSE characteristics like CFO to guide model adaptation. Unlike conventional approaches, our core innovation lies in using physically-grounded signal impairment transformations to generate reliable prototype points, which then provide well-directed adaptation targets. This physics-aware design specifically addresses the unique challenges of RFF distribution shifts on DSE. Our main contributions are as follows:

- To the best of our knowledge, we are the first to propose a *RFF-TTA* technique that combines physical impairments to drive model adaptation. Our *RFF-TTA* can significantly improve the existing RFF method’s performance upper bound.
- We propose a physical impairment-based (CFO and CPO) data augmentation framework targeted to overcome the DSE by systematically characterizing the random drift properties of hardware impairments.
- We propose a physical information-aware prototype ensemble approach by utilizing physical impairments-

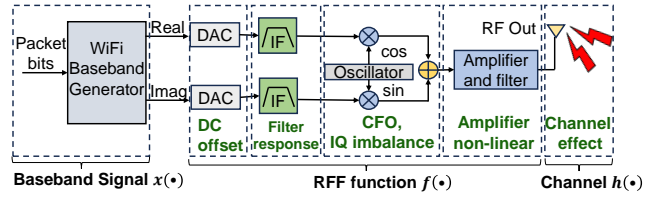


Figure 3: Typical digital orthogonal modulation transmitters of WiFi.

based augmented data. The method is non-parametric-based and can directly utilize unlabeled online batch flow data to implement TTA.

Background and Challenges

Temporally-Varying RF Fingerprint Modeling

This study examines RFF identification in WiFi-enabled mobile devices using IEEE 802.11 OFDM systems. During signal transmission, as shown in Fig. 3, various analog components (including DACs, IF filters, oscillators, power amplifiers, and antennas) imprint unique hardware-specific characteristics onto the signals. These components introduce distinct physical-layer impairments such as DC offset, CFO, and I/Q imbalance, which collectively form distinctive device signatures (Huang et al. 2017). We mathematically model these hardware imperfections through a composite function $f(\cdot)$. For a given baseband signal $x_b(t)$ at time t , its transformed version after traversing the RF front-end can be represented as $f(x_b(t))$. Furthermore, to account for the electromagnetic channel’s influence during wireless propagation, we introduce a channel response function $h(\cdot)$. Consequently, the received signal $r_k(t)$ from the k -th transmitting device can be mathematically expressed as

$$r_k(t) = f_k(x_b(t)) * h(t) + wgn(t), k \in \{1, 2, \dots, K\}, \quad (1)$$

where the K denotes the number of transmitters, $*$ denotes the linear convolution operation, $wgn(t)$ denotes the additive white Gaussian noise (AWGN).

In real-world RFF implementations (Fig. 1), dynamic device workloads and fluctuating channel conditions can significantly alter fingerprint characteristics. Our analysis begins with the established hardware imperfection framework from (Zheng, Sun, and Ren 2019), which provides fundamental insights into these variations. Specifically, we examine

$$f_k(x_b(t), \tau) = A(1 + pow^k(x_b(t), \tau)) \cdot e^{i(w_c + \Delta w^k + \Theta)2\pi(nT_s + \tau) + i\varphi^k(x_b(t), \tau)} \quad (2)$$

where A denotes the amplitude, nT_s is the sampling interval, τ is the sampling phase, w_c is the carrier frequency, Δw^k denotes the CFO of k -th device, $\Theta(nT_s + \tau)$ is modulation information, $pow^k(x_b(t), \tau)$ and $\varphi^k(x_b(t), \tau)$ denote the nonlinear part of the amplitude and CPO, respectively.

Since signal propagation goes through wireless channels, we apply channel effects $h(\cdot)$ to the transmit signal in Eq. 2, so that the signal at the receiver in Eq. 1 is simplified as

$$r_k(x_b(t), \tau) = f_k(x_b(t), \tau) e^{i(2\pi\Delta w^k + \varphi^k)(x_b(t), \tau)} * h(t) + wgn(t). \quad (3)$$

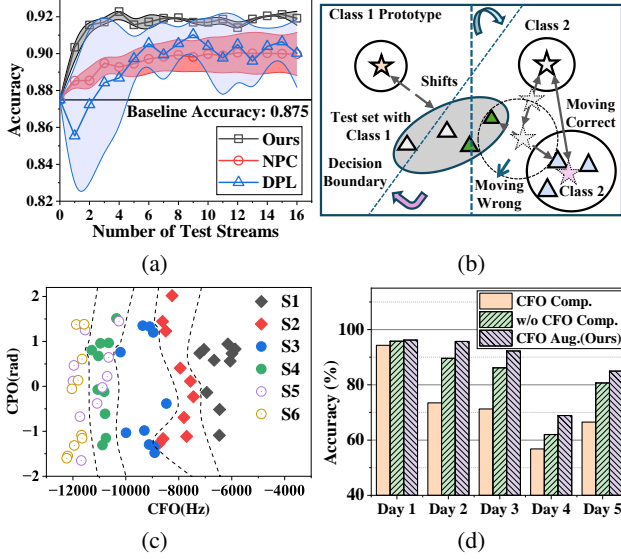


Figure 4: (a) Comparative analysis of different TTA algorithms for RFF; (b) Analysis of the causes of TTA failure; (c) RF physical feature distribution offset in different days, devices and application loads, S1~S2: Day₁-Dev₀-App_{0~1}, S3~S4: Day₁-Dev₁-App_{0~1}, S5~S6: Day₂-Dev₁-App_{0~1}; (d) Effectiveness analysis of CFO compensation.

Online TTA Definition

To address such distribution shifts, TTA offers an efficient solution by leveraging target domain batch data for real-time model adaptation, eliminating the need for retraining while maintaining low computational overhead—making it particularly suitable for online testing scenarios. Assuming the existence of multi-source domain dataset denoted as $\mathcal{D}_{train} = \{D_s\}_{s=1}^S$ with $D_s = \{r_s(t) = (x_s^i, y_s^i)\}_{i=1}^N$, where $x_s^i \in \mathbb{C}^M$ denotes a signal sample with dimension M in complex space \mathbb{C} , $y_s^i \in \mathbb{R}^K$ denotes the K -dimensional label by one-hot encoding in real number space \mathbb{R} , S and N denote the number of domains and the number of samples, respectively. Similarly, we define the T target domain test sample as $\mathcal{D}_{test} = \{D_t\}_{t=0}^T$ with $D_t = \{r_t = (x_t^i, y_t^i)\}_{i=1}^N$. It is noteworthy that $D_S \neq D_T$, and probability distribution shift exists between $p(x_S)$ and $p(x_T)$. Specifically, we first train a DL model $F_\theta : x_i \rightarrow y_i$ on D_s . Then, for a stream of unlabeled batch test samples $\{x_b \sim p(x_T)\}_{b=0}^B$, TTA aims to minimize the deviation of the distribution of model features $F_\theta(y|x)$ from the distribution of test data $P(y_b|x_b)$, i.e.,

$$\min_{\theta} E_{x_b \sim p(x_T)} \|F_\theta(y|x) - P(y_b|x_b)\| \rightarrow 0 \quad (4)$$

where b denotes the batch index, $\|\cdot\|$ denotes the distance.

Analysis of TTA Challenges under Time-Varying RFF

To realize online TTA for RFF, the most mainstream approach is currently based on class prototype movement learning, such as DPL(Wang et al. 2025) and NPC(Zhang

et al. 2023), which has a minimal computational overhead and does not require model optimization training. Fig. 4 (a) shows the results of these TTA methods for RFF. We can observe that these methods are unstable compared to ours, which have larger error bands, and the DPL also produces an accuracy degradation compared to baseline accuracy.

Instability Analysis. As shown in Fig. 4 (b), the feature space contains distinct prototypes for class 1 (P_1) and class 2 (P_2), with corresponding batch test samples $\{x_1^b\}_{i=0}^B$ and $\{x_2^b\}_{i=0}^B$. There is a vertical decision boundary line in the space dividing the space into left and right parts. When testing adaptive, few-shot test samples $\{x_1^b\}_{i=0}^B$ of class 1 are assigned pseudo-labels as y_2 due to domain bias (red triangles in the figure). Due to the noise in the pseudo-labeling of the test sample, the P_2 will move towards that sample $\{x_1^b\}_{i=0}^B$, which results in more test samples of x_1 being misclassified as y_2 . The wrong direction of prototype movement will result in model performance degradation. Therefore, *this paper focuses on how to design reliable test adaptive directions for RFF recognition models.*

Solution Ideas. Based on the above analysis, if we increase the sample size of class 1 and class 2, a more realistic distribution will be revealed, thus providing more accurate prototype points for the target domain. However, in the online test stream, the samples are reached in batches, and there is a problem of insufficient sample size. To increase the sample size, an intuitive approach is data augmentation. However, most of the data augmentation in TTA is image-based, which makes it difficult to apply to RFF. For this reason, we need to explore the key factors of the time-varying RFF distribution bias and use them to design a suitable augmentation strategy.

Fortunately, as described earlier in Eq. 3, the factors of RFF offset mainly contain CFO and CPO. Fig. 4 (c) illustrates the phenomenon of CFO and CPO shift of RFF for different time and application load cases. Conventional wisdom might suggest applying CFO compensation techniques, a common preprocessing step in many RFF systems (Tang et al. 2024), to eliminate this instability. However, our empirical results, presented in Fig. 4 (d), reveal a counter-intuitive finding: applying standard CFO compensation dramatically degrades recognition accuracy in our test scenario. This contradicts previous work, which often relied on dedicated WiFi modules with stable thermal profiles and thus stable RFF signatures. In contrast, for mobile devices with complex, fluctuating workloads, the CFO drift is not merely noise but a rich, state-dependent feature. *The above observations provide important data augmentation ideas for our TTA implementation, i.e., CFO and CPO augmentation can generate samples to approximate the true distribution, and thus determine a more accurate direction of adaptation for prototype-based TTA.*

Physical Information-Aware Prototype-based RFF-TTA scheme

Overview

To enhance the DG capability of radio frequency fingerprint (RFF) recognition, we propose RFF-TTA, a novel

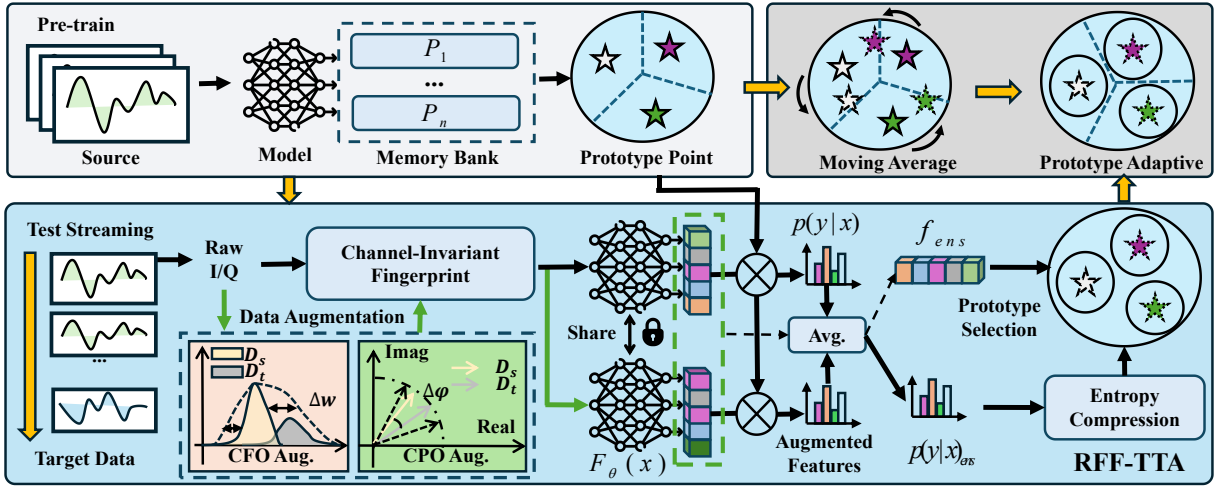


Figure 5: Framework diagram of the RFF-TTA. The model employs prototype learning for pre-training to obtain category prototypes. During testing, each data stream undergoes CFO/CPO augmentation before feature integration. An entropy compression scheme then computes test prototypes, which adaptively update memory bank prototypes via sliding average.

TTA method leveraging physics-aware prototypes. Our design idea is mainly to utilize CFO and CPO augmentation, which are physical impairments, to get a more accurate sample distribution. From there, we determine the physics-aware points of the test samples and guide the model prototype points for sliding adaptation. Specifically, as illustrated in Fig. 5, our approach first pre-trains source-domain data to establish category prototypes stored in a memory bank. During online inference, we perform physics-informed augmentation on streaming test data, extract and fuse features from both original and augmented samples. Further, we integrate the two deep features and classification probabilities. Then we employ entropy-based confidence selection to derive target-domain prototypes. These are progressively integrated into the memory bank via exponential moving average, enabling continuous model adaptation to new domains.

Feature Pre-training Stage

To overcome channel-induced distribution shifts in RF fingerprinting, we develop a dual approach combining channel-invariant feature extraction with prototype learning.

RF Signal Introduction. In this paper, we mainly use device information-independent WiFi preamble signals to perform the RFF recognition task. The IEEE 802.11 OFDM standard defines a preamble at the beginning of the packet, which consists of 10 short training symbols (STS), 1 guard interval (GI), and 2 long training symbols (LTS) (Xing et al. 2022). For the received RF signals $x(t)$, we first use a synchronization technique (Liu et al. 2019) to determine the starting point of the signal, and then use root-mean-square normalization.

Channel-Invariant Fingerprint. Since different symbols in the same RF signal go through the same channel, we mainly utilize different adjacent symbol bits in the preamble signal, i.e., STS and LTS, to cancel the channel (Xing et al. 2022). Our main approach is to use the Fast Fourier Transform (FFT) and logarithmic operations \mathcal{F}_{\log} to transform the

channel time-domain convolution in Eq. 3 into a frequency-domain addition that eliminates the channel. For more information, please refer to the (Xing et al. 2022) and Appendix. Our channel-invariant RFF consists of two main components, namely Steady RFF Symbol Pairs and Transient RFF Symbol Pairs.

Steady RFF Symbol Pairs. In the preamble, the first STS signal, as well as the GI, usually contains transient information for device startup and information transition. In contrast, the 2-th to 10-th STS as well as LTSs are steady-state signals. We use the 2nd to 5th STS $x_{2:5}^{STS}$ to compute the logarithmic spectrum and then perform a subtraction operation with the first LTS x_1^{LTS} to eliminate the channel. Therefore, we define the steady-state RFF as

$$RFF_{Stable} = \{(\mathcal{F}_{\log}(x_{2:5}^{STS}) - \mathcal{F}_{\log}(x_1^{LTS})), (\mathcal{F}_{\log}(x_{5:9}^{STS}) - \mathcal{F}_{\log}(x_2^{LTS}))\}, \quad (5)$$

where $x_{5:9}^{STS}$ denotes the 2nd to 5th STSs, x_2^{LTS} denotes the 2-th LTS.

Transient RFF Symbol Pairs. Similarly, we define the remaining transient RFF to be

$$RFF_{Transient} = \{F_{\log}(x_{0:4}^{STS}) - F_{\log}(\text{concat}(x_{9:10}^{STS}, x^{GI}, x_{1*}^{LTS}))\}, \quad (6)$$

where x_{1*}^{LTS} denotes the first 16 bits of the 1-th LTS, x^{GI} denotes GI segment, $x_{9:10}^{STS}$ denotes the 9-th to 10-th STSs, $\text{concat}(\cdot)$ denotes the concatenate operation.

Prototype-based model pre-training. After that, we integrate $RFF_{Transient}$ and RFF_{Stable} to \tilde{x} and use ResNet-18 (He et al. 2016) as a feature extractor to perform the classification task. Specifically, we first train the feature embedding network $F_{\theta}(\bullet)$ and classifier W_0 . The classifier weights w_i are represented by the average embedding of each new class y_i (i.e., the class prototype or the most representative feature of the class). The class prototype w_j in D_s can be

calculated by:

$$w_j = \left\| \frac{1}{|D_s|} \sum_{i=1}^{|D_s|} I(y_i = j) F_\theta(\tilde{x}_i) \right\|_2, \quad (7)$$

where $\|\bullet\|_2$ denotes the l_2 normalization and $I(\cdot)$ denotes the indicator function. With the class prototype points w_i , we can calculate the class probability p_j for each train sample as

$$p(y = j|\tilde{x}; F_\theta) = \frac{\exp(\text{sim}(F_\theta(\tilde{x}), w_j))}{\sum_{j \in y_s} \exp(\text{sim}(F_\theta(\tilde{x}), w_j))}, \quad (8)$$

where $\text{sim}(\bullet)$ denotes the cosine similarity function. The Eq. 8 suggests that the similarity of the sample to the class prototype point determines the sample prediction category.

Finally, we use prototype contrastive learning loss ℓ_{contr} for pre-training:

$$\ell_{contr} = -\frac{1}{|D_s|} \sum_{i=1}^{|D_s|} \log\left(\frac{\exp(\text{sim}(F_\theta(\tilde{x}), w_{y_i}))}{\sum_{j \in y_s} \exp(\text{sim}(F_\theta(\tilde{x}), w_j))}\right). \quad (9)$$

Using Eq. 9 to train the DL model, we can get the prototype points w_j for each category y_j , which are saved in the memory bank as $P_s = \{w_0, w_1, \dots, w_j | w_j \in D_s\}$.

Online Test Time Adaptive Scheme for RFF

Physical Impairment Data Augmentation. We aim to design TTA strategies using batch online test data streams to improve the generalization of DL models to unseen RFF distributions. We first utilize the physical information (CFO, CPO) for data augmentation and then compute the prototype point as an adaptive direction.

CFO Random Augmentation. To determine the scope of the augmentation, we first use the CFO estimation method in (Shen et al. 2021), which consists mainly of coarse frequency offset $\Delta\hat{w}^{coarse}$ estimation and fine frequency offset $\Delta\hat{w}^{fine}$ estimation. Then, we perform frequency random offset augmentation on the source domain. We define a variable α obeying a uniform distribution: $\alpha \sim U[0, 1]$. We design the randomly augmented frequency offset to be $\alpha\Delta f \sim U[0, \Delta f]$, where $\Delta f \in \max[\Delta\hat{w}^{fine}, \Delta\hat{w}^{coarse}]$ denotes the frequency offset range. In combination with Eq. 3 and without considering noise, the frequency augmentation can be calculated by the following equation

$$\hat{x}(t)^f = f_k(x_b(t), \tau) e^{i(2\pi(\Delta w^k + \alpha\Delta f) + \varphi^t)(x_b(t), \tau)} * h(t). \quad (10)$$

CPO Random Augmentation. Similarly, we propose random phase $\Delta\varphi$ augmentation by utilizing a random variable $\beta \sim U[-1, 1]$ to control. The augmented phase offset can be denoted as $\beta\Delta\varphi \sim U[-\beta\Delta\varphi, \beta\Delta\varphi]$. In combination with Eq. 3, the phase augmentation can be calculated by the following equation

$$\hat{x}(t)^p = f_k(x_b(t), \tau) e^{i(2\pi\Delta w^k + \varphi^t + \beta\Delta\varphi)(x_b(t), \tau)} * h(t). \quad (11)$$

We perform channel-invariant fingerprint extraction of the augmented $\hat{x}(t)^f$ and $\hat{x}(t)^p$ through Eqs. 5 and 6, which are then fed into the DL model to obtain the augmented features $F_\theta(\tilde{x}_{aug})$.

Physics Informed-Aware Prototypical Ensemble. For on-line batch flow data, we use the original test data \tilde{x}_{raw} and its augmented sample pairs \tilde{x}_{aug} to jointly perceive the prototype point distribution of the test set. We first average this sample pairwise and categorical probabilities to obtain the ensemble features $F_\theta(\tilde{x})_{ens}$ and probabilities $p(\tilde{x})_{ens}$, respectively, as

$$\begin{aligned} F_\theta(\tilde{x})_{ens} &= (F_\theta(\tilde{x}_{raw}) + F_\theta(\tilde{x}_{aug}))/2 \\ p(\tilde{x})_{ens} &= (p(\tilde{x})_{raw} + p(\tilde{x})_{aug})/2 \end{aligned} \quad (12)$$

where $F_\theta(\tilde{x}_{raw})$ denote the extracted original feature, $p(\tilde{x})_{raw}$ and $p(\tilde{x})_{aug}$ denote the probability values for the original sample classification and the augmented sample, respectively, which can be calculated using Eq. 8.

We then perform the computation of the physical-aware prototype P_{ens} . We can assign pseudo-labels \tilde{y} for each sample by $p(\tilde{x})_{ens}$, i.e., $\tilde{y} = \arg \max p(\tilde{x})_{ens}$. Based on the pseudo-labeling, we can classify the test samples into categories so that we can calculate the prototype points of each test category. However, since pseudo-labeling contains noise, i.e., false predictions, this can lead to bias in the calculation of physically perceived prototype points. Therefore, we need to filter the test samples and extract high-confidence samples as the source of prototype calculation samples. We mainly use entropy compression methods to extract high-confidence samples, which can be computed by

$$H_{ens}^i = -\sum_{j=1}^{y_s} \sigma(p(\tilde{x})_{ens}^j) \log \sigma(p(\tilde{x})_{ens}^j) \quad (13)$$

where H_{ens}^i denotes information entropy of the i -th x_b^i , σ denotes the softmax function. For the batch test sample information entropy set $\{H_{ens}^i\}_{i=0}^N$, we select top_K samples,

$$\begin{aligned} top_K &= \arg \text{sort} \{H_{ens}^i\}_{i \in \{1, \dots, N\}} \\ S_K &= \{x_b^i | I(i \in top_K) = 1\} \end{aligned} \quad (14)$$

where $\text{sort}(\cdot)$ denotes the sorting function, S_K denotes the subset of samples that are selected. Eq. 14 clarifies that the smallest information entropy samples ranked as K are selected, and these samples usually have higher classification confidence.

Finally, using S_K , we can compute the physical-aware prototype on test set as

$$\begin{aligned} P_{ens} &= \{\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_j | \tilde{w}_j \in S_K\}, \\ \tilde{w}_j &= \left\| \frac{1}{|S_K|} \sum_{i=1}^{|S_K|} I(y_i = j) F_\theta(\tilde{x})_{ens} \right\|_2. \end{aligned} \quad (15)$$

Exponential sliding average of prototype points.. Once the physically-aware prototypes P_{ens} of the test set are available, we will move the prototype points P_s in the memory bank so that the model gradually adapts to the class distribution of the test data. Since the test data are reached in batches, for the t -th arrival, the prototype points P_M^t in its

Table 1: Identification Accuracy of ResNet with Different Feature Representations in Single-Source Domain.

Source	Target	Classification Accuracy (\uparrow)											
		<i>RFF-TTA</i>	CLPS	DoLos	ChIns	FFT	IQ	PSD	EPS	STST	CWT	HHT	EMD
E1_D1	E1_D1	0.9630	0.9750	0.9250	0.9125	0.9141	0.9297	0.9688	0.9219	0.9375	0.9453	0.9366	0.9312
	E1_D2	0.9118	0.8750	0.8625	0.8249	0.7891	0.8125	0.8203	0.7969	0.8516	0.8672	0.7656	0.7734
	E1_D3	0.9521	0.7624	0.8625	0.8125	0.7266	0.8047	0.7188	0.7134	0.8359	0.7422	0.7500	0.7578
	E1_D4	0.8091	0.6125	0.4875	0.7125	0.4297	0.4368	0.4844	0.4531	0.5859	0.5836	0.4688	0.5156
	E1_D5	0.9123	0.7750	0.7750	0.8125	0.7656	0.7734	0.7500	0.7322	0.8125	0.8234	0.6953	0.7031

memory bank can be updated simultaneously with the following equation

$$\begin{cases} P_M^t = \gamma P_M^{t-1} + (1 - \gamma) P_{ens}^t, t \geq 1 \\ P_M^t = P_s, t = 0 \end{cases} \quad (16)$$

where $\gamma \in [0, 1]$ denotes the momentum coefficient. Using Eq. 16, our model can continuously estimate P_{ens}^t based on the received data, followed by a gradual migration of the class distribution centered on the prototype points, leading to online TTA.

Classification. Finally, we mainly use the updated prototype points P_M^t and Eq. 8 for classification probability calculation,

$$p(y = j | \tilde{x}; F_\theta) = \frac{\exp(\text{sim}(F_\theta(\tilde{x}), P_M^{tj}))}{\sum_{j \in y_s} \exp(\text{sim}(F_\theta(\tilde{x}), P_M^{tj}))}. \quad (17)$$

Experiments and Analysis

Dataset. To validate the effectiveness of *RFF-TTA*, we collected time-varying RFF datasets from the real world. We used 10 mobile phones with complex tasks as signal transmitters. Because running different applications on the phone produces different DSEs and thus time-varying RFFs, we then conducted tests at different times and locations to simulate the impact of channel effects. The RF signal from all the devices is captured using the USRP N210 software radio (SDR) with a sampling rate of 20 MS/s. 10,000 samples per day and 2,000 samples per device were used for testing. In total, we collected data for 9 days and 2 months. For more information, please refer to the Appendix.

Experiments setup

Hardware Setup. The *RFF-TTA* algorithm is implemented by using Pytorch 1.10.0 and executed on a computer running Ubuntu 18.04.6 LTS, with Intel(R) Core(TM) i9-10900K CPU@3.70 GHz and 2 NVIDIA GeForce RTX3090 GPUs.

Models' Setup and Comparison Baselines. To ensure the fairness of the comparison, we use a uniform network architecture (ResNet-18 (He et al. 2016)) for all baseline methods. The state-of-the-art RFF recognition algorithms we compare include CLPS (Tang et al. 2024), DoLos (Xing et al. 2022), ChIns (Shen et al. 2022). In addition to this, other signal features are considered for comparison, including power spectral density (PSD), envelope's power spectrum (EPS)(Elmaghub and Hamdaoui 2023), short-time Fourier transform (STST), continuous wavelet transform (CWT), Hilbert Huang transform (HHT), and empir-

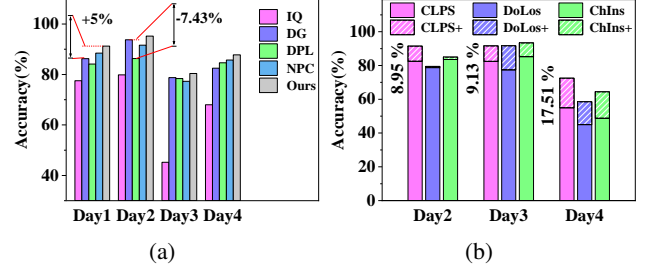


Figure 6: (a) Performance comparison of different TTA algorithms. (b) Efficiency of RFF-TTA imposed on other RFF recognition algorithms.

ical modal decomposition (EMD). In addition, we considered two advanced TTAs (NPC (Zhang et al. 2023) and DPC (Wang et al. 2025)) for comparison. Both algorithms are used for real-time TTA and are more similar to our scenario.

Metrics. To quantitatively evaluate the RFF recognition performance, we use Accuracy (Acc) as a metrics

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (18)$$

where TP, TN, FP, FN denote the true-positives, true-negatives, false-positives, and false-negatives, respectively.

Implementation Details. We set Δf to 2000 MHz for CFO augmentation and to $[-2\pi, 2\pi]$ in CPO augmentation. K in Eq. 14 is set to 5 and the parameter γ in Eq. 16 is set to 0.5. The model optimization is mainly done using the Adam optimizer with a learning rate of 0.001.

Evaluating the Generalization of *RFF-TTA*

Comparison of different RF fingerprint recognition algorithms. We use the data collected on the first day for model training and then test it in the same environment on the following days. Table 1 demonstrates that our *RFF-TTA* improves the average recognition performance by 8.7% compared to the state-of-the-art RFF recognition algorithm (CLPS). In addition, we find that RFF is slightly lower than CLPS in the test of E1_D1, while CLPS performs poorly in the later tests. This suggests that CLPS produces overfitting on the same distributed data, whereas our method mitigates the overfitting well and thus improves the generalization. Fig. 7a illustrates the prototype point adaptation results for *RFF-TTA*. We can see that the method in this paper can better move the prototype points with distribution offset to the target domain data. The above results show that the *RFF-TTA* in this paper can better cope with channel effects and DSE.

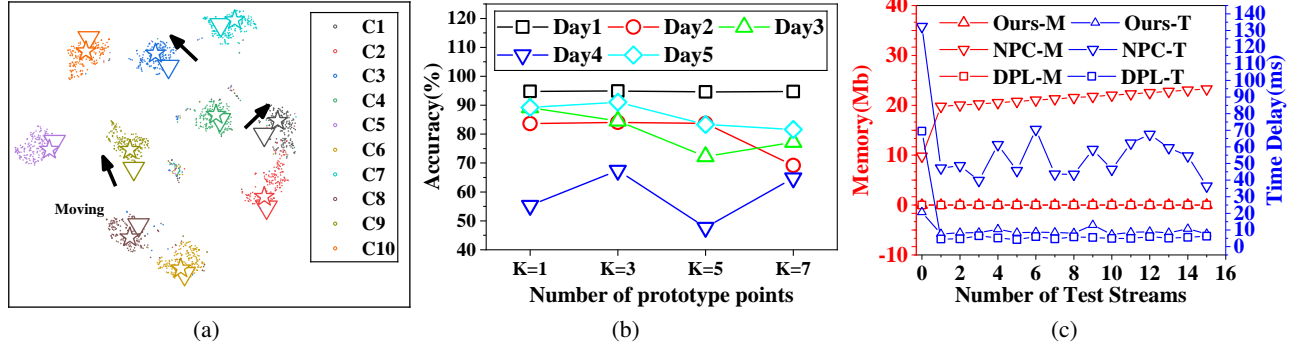


Figure 7: (a) The t-SNE diagram of the updated prototype point is passed through the *RFF-TTA*, ∇ and \star denote the old prototype and the adapted prototype, respectively. (b) Sensitivity analysis of the sample selection parameter K . (c) Running latency and memory consumption of different TTA algorithms.

Table 2: Comparing the RFF recognition efficiency of different training modules: M1 denotes the stable features, M2 denotes the CFO augmentation, M3 denotes the CPO augmentation, and M4 denotes the transient features.

M1	M2	M3	M4	Day6	Day7	Day8	Day9
-	-	-	-	0.8101	0.7698	0.6010	0.6421
✓	-	-	-	0.9153	0.8762	0.8257	0.8684
✓	✓	-	-	0.9195	0.8952	0.8563	0.8934
✓	✓	✓	-	0.9512	0.9173	0.8624	0.8944
✓	✓	✓	✓	0.9674	0.9426	0.8878	0.9355

Comparison of different TTA algorithms. Since the study in this paper focuses on online TTA, we mainly compare it with two more advanced online TTA methods (NPC, DPL). Figure 7a illustrates the results, and it can be seen that our RFF has better performance improvement compared to DG, where DG denotes our channel-invariant fingerprint method, and IQ denotes the raw IQ signal. The performance of DPL as well as NPC is not stable due to their inability to adapt to the RFF data characteristics and find the exact direction of prototype update.

Effectiveness of the TTA algorithm for other RFF extraction methods. Fig. 6b shows the efficiency of the TTA strategy proposed in this paper imposed on other RFF algorithms. It can be seen that the *RFF-TTA* can improve the recognition performance of other algorithms by an average magnitude of about 8%. The above results demonstrate the applicability of the *RFF-TTA*, which can be effectively embedded into other methods and improve their performance upper bound.

Ablation Study

Analysis of the validity of data augmentation. Table 2 demonstrates the performance differences resulting from different data augmentation (Iwana and Uchida 2021) and RFF features. The channel-invariant fingerprinting adopted in this paper can eliminate the channel effect, thereby improving performance. In addition, the CFO augmentation and CPO augmentation used in this paper can increase data diversity and also improve the model’s generalization performance, thus mitigating the effect of DSE. Additionally,

Table 3: Performance comparison of different time series augmentation methods.

	Ours	Gaussian noise	Channel Effect	Scaling	Amplitude distortion	Time Warp
Day1	0.9630	0.9695	0.9673	0.9517	0.9599	0.9492
Day2	0.9118	0.8465	0.7691	0.8414	0.8444	0.8398
Day3	0.9475	0.8578	0.8200	0.8346	0.8127	0.8330

we compare various time series augmentation methods. As shown in Table 3, it can be seen that our data augmentation achieves the best results, indicating that our approach is more aligned with the DSE problem explored in this paper.

Sensitivity of the sample selection parameter K . Eq. 14 requires the selection of Top- K samples for target adaptation. Figure 7b shows the results for different values of K . It can be seen that a larger value of K is not better. Choosing a smaller K value indicates a higher confidence in the sample.

Operational efficiency of different TTA algorithms.. Fig. 7c illustrates the operational efficiency of different TTA algorithms. We can see that NPC has the highest time delay and memory consumption because it needs to keep adding features to the memory bank. *RFF-TTA* has the same memory consumption as DPL, and the prototype memory bank does not grow because only prototype updates are implemented instead of adding new prototype points. However, RFF computational latency is higher than DPL, but still within acceptable limits. In contrast. Finally, in terms of performance, as shown in Fig. 6a, ours is the best and DPL the worst.

Conclusion

In this paper, we propose a physical impairment-based augmented method and a physical information-aware prototype-based TTA method to solve the DG problem in time-varying RFF. We first propose random CFOs and CPOs to augment the data. Then the combination of transient and steady-state RFF is used to eliminate the channel effect. Finally, we propose a physical information-aware prototype ensemble approach to implement TTA. Our approach has achieved initial results, but extensive application testing in larger IoT scenarios is needed in the future.

References

- Brik, V.; Banerjee, S.; Gruteser, M.; and Oh, S. 2008. Wireless device identification with radiometric signatures. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*, 116–127.
- Cai, Z.; Lu, G.; Wang, Y.; Gui, G.; and Sha, J. 2025. Robust Cross-Domain UAV RFFI Method Using Domain-Invariant Adversarial Learning and Manifold Regularization. *IEEE Transactions on Cognitive Communications and Networking*.
- Elmaghub, A.; and Hamdaoui, B. 2023. A Needle in A Haystack: Distinguishable Deep Neural Network Features for Domain-Agnostic Device Fingerprinting. In *2023 IEEE Conference on Communications and Network Security (CNS)*, 1–9. IEEE.
- Gong, P.; Ragab, M.; Wu, M.; Chen, Z.; Su, Y.; Li, X.; and Zhang, D. 2025. Augmented contrastive clustering with uncertainty-aware prototyping for time series test time adaptation. *arXiv preprint arXiv:2501.01472*.
- Gu, X.; Wu, W.; Song, A.; Yang, M.; Ling, Z.; and Luo, J. 2024. RF-TESI: Radio Frequency Fingerprint-based Smartphone Identification under Temperature Variation. *ACM Transactions on Sensor Networks*, 20(2): 1–21.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Huang, Y.; et al. 2017. Radio frequency fingerprint extraction of radio emitter based on I/Q imbalance. *Procedia computer science*, 107: 472–477.
- Iwana, B. K.; and Uchida, S. 2021. An empirical survey of data augmentation for time series classification with neural networks. *Plos one*, 16(7): e0254841.
- Jana, S.; and Kasera, S. K. 2008. On fast and accurate detection of unauthorized wireless access points using clock skews. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*, 104–115.
- Jiang, Q.; Ye, C.; Wei, D.; Xue, Y.; Jiang, J.; and Wang, Z. 2024. Discover Your Neighbors: Advanced Stable Test-Time Adaptation in Dynamic World. *arXiv preprint arXiv:2406.05413*.
- Li, Y.; Lin, H.; Lv, J.; Gao, Y.; and Dong, W. 2024. BLE Location Tracking Attacks by Exploiting Frequency Synthesizer Imperfection. In *IEEE INFOCOM 2024-IEEE Conference on Computer Communications*.
- Liu, P.; Yang, P.; Song, W.-Z.; Yan, Y.; and Li, X.-Y. 2019. Real-time identification of rogue WiFi connections using environment-independent physical features. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, 190–198. IEEE.
- Polak, A. C.; and Goeckel, D. L. 2015. Identification of wireless devices of users who actively fake their RF fingerprints with artificial data distortion. *IEEE Transactions on Wireless Communications*, 14(11): 5889–5899.
- Rajendran, S.; and Sun, Z. 2022. RF impairment model-based IoT physical-layer identification for enhanced domain generalization. *IEEE Transactions on Information Forensics and Security*, 17: 1285–1299.
- Rajendran, S.; Sun, Z.; Lin, F.; and Ren, K. 2020. Injecting reliable radio frequency fingerprints using metasurface for the Internet of Things. *IEEE Transactions on Information Forensics and Security*, 16: 1896–1911.
- Restuccia, F.; D’Oro, S.; Al-Shawabka, A.; Belgiovine, M.; Angioloni, L.; Ioannidis, S.; Chowdhury, K.; and Melodia, T. 2019. DeepRadiolD: Real-time channel-resilient optimization of deep learning-based radio fingerprinting algorithms. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 51–60.
- Riyaz, S.; Sankhe, K.; Ioannidis, S.; and Chowdhury, K. 2018. Deep learning convolutional neural networks for radio identification. *IEEE Communications Magazine*, 56(9): 146–152.
- Shen, G.; Zhang, J.; Marshall, A.; and Cavallaro, J. R. 2022. Towards scalable and channel-robust radio frequency fingerprint identification for LoRa. *IEEE Transactions on Information Forensics and Security*, 17: 774–787.
- Shen, G.; Zhang, J.; Marshall, A.; Peng, L.; and Wang, X. 2021. Radio frequency fingerprint identification for LoRa using spectrogram and CNN. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, 1–10. IEEE.
- Tang, P.; Xu, Y.; Ding, G.; Jiao, Y.; Song, Y.; and Wei, G. 2024. Causal Learning for Robust Specific Emitter Identification over Unknown Channel Statistics. *IEEE Transactions on Information Forensics and Security*.
- Wang, G.; Ding, C.; Tan, W.; and Tan, M. 2025. Decoupled Prototype Learning for Reliable Test-Time Adaptation. *IEEE Transactions on Multimedia*.
- Xing, Y.; Hu, A.; Zhang, J.; Peng, L.; and Wang, X. 2022. Design of a channel robust radio frequency fingerprint identification scheme. *IEEE Internet of Things Journal*, 10(8): 6946–6959.
- Zhang, M.; Tang, P.; Wei, G.; Ni, X.; Ding, G.; and Wang, H. 2024. Open set domain adaptation for automatic modulation classification in dynamic communication environments. *IEEE Transactions on Cognitive Communications and Networking*.
- Zhang, Y.; Wang, X.; Jin, K.; Yuan, K.; Zhang, Z.; Wang, L.; Jin, R.; and Tan, T. 2023. Adanpc: Exploring non-parametric classifier for test-time adaptation. In *International conference on machine learning*, 41647–41676. PMLR.
- Zheng, T.; Sun, Z.; and Ren, K. 2019. FID: Function modeling-based data-independent and channel-robust physical-layer identification. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, 199–207. IEEE.