Towards Collaborative and Cross-Environment UAV Classification: Federated Semantic Regularization

Xue Fu, *Graduate Student Member, IEEE*, Yu Wang, *Member, IEEE*, Yun Lin, *Senior Member, IEEE*, Tomoaki Ohtsuki, *Senior Member, IEEE*, Bamidele Adebisi, *Senior Member, IEEE*, Guan Gui, *Fellow, IEEE*, and Hikmet Sari, *Life Fellow, IEEE*

***Abstract*—The rapid and widespread adoption of unmanned aerial vehicles (UAVs) poses significant threats to public safety and security in sensitive areas and subsequently underscores the urgent need for effective UAV surveillance solutions, where UAV classification emerges as a vital technology. Deep learning (DL) methods can autonomously extract implicit features from UAV signals and subsequently infer their types, provided that sufficient signal samples are available. Due to the high mobility of UAVs, it is challenging to ensure continuous monitoring between UAVs and the surveillance system to obtain sufficient samples. Moreover, DL models developed from sufficient but environment-specific datasets tend to be less generalized. This paper proposes a novel federated semantic regularization for learning an UAV classification model and further classifying UAVs across diverse environmental conditions. The approach enhances model generalization by regularizing semantic features during the local model training process on each participant. Subsequently, these local models are aggregated into a robust global model. Extensive testing across multiple environments demonstrates the superior classification performance of our approach compared to existing non-federated and federated approaches. The average classification accuracy of the proposed method in the three environments is 95.68%, which is improved by 13.39% compared to the non-federated methods and by 2.75% compared to the federated methods.**

***Index Terms*—UAV classification, federated learning, prototype learning.**

I. Introduction

Unmanned aerial vehicles (UAVs) play an important role in various services such as goods delivery, disaster monitoring or military and have been regard as a key technology in 5G/6G wireless networks due to its high flexibility and seamless connectivity [1], [2]. Nevertheless, the rapid and widespread usage of UAV technology poses substantial threats to public safety and the potential compromise of security in sensitive zones. Regulations do not allow UAVs to fly over all areas, in addition to typical altitude limitations, but these regulations are easy to crack, resulting in lifting the limits of height and no-fly zone. As a result, there is an urgent need for UAV surveillance solutions.

Radio frequency (RF) sensing is one of the key technologies used in surveillance solutions to classify UAV. The relevant research on RF-based UAV classification mainly falls into two categories, namely handcrafted feature-based and deep learning (DL)-based approaches. Handcrafted features [3], [4] are typically derived from observations or calculations based on captured signals and generally depend on expert knowledge. In contrast, DL methods can autonomously extract implicit features from UAV signals and subsequently infer their types, provided that sufficient signal samples are available. This autonomous feature extraction by DL models not only reduces the reliance on specialized expertise but also enhances the adaptability and accuracy of the classification processes [5]–[7]. However, due to the high mobility of UAVs, it is challenging to ensure the continuous listening between UAVs and a surveillance system. In such cases, signal samples captured by the surveillance system are not sufficient to drive the training process of the DL model. Even if the surveillance system can capture sufficient samples, the DL model developed from sufficient but environment-specific samples tends to be less generalized. Integrating the samples from multiple surveillance systems through wireless transmission could address this challenge, but it is infeasible or impractical due to limited communication resources, data privacy concerns, or country regulations [8].

Federated learning (FL) [9] can offer various important benefits such as data privacy enhancement, low-latency network communication and enhanced learning quality for DL applications [2]. With the help of FL, UAV surveillance systems can obtain a much better level of classification by coordinating multiple systems to perform training process while keeping data safe. Due to the high mobility of UAVs and the different locations of surveillance systems, there are significant distribution differences between the signal samples of UAVs captured by multiple surveillance systems. Such distribution differences make surveillance systems suffer from learning-drift, resulting in unstable and slow convergence.

This work was supported in part by the Natural Science Foundation of China under Grant 62471247, and Grant 62401281, in part by the Key Project of the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province under Grant 22KJA510002, in part by Japan Science and Technology Agency (JST) Adopting Sustainable Partnerships for Innovative Research Ecosystem (ASPIRE) program under Grant JPMJAP2326, and the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX23_1024. (*Corresponding authors: Guan Gui, Yun Lin*).

Xue Fu, Yu Wang, Guan Gui, and Hikmet Sari are with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: 1020010415@njupt.edu.cn, yuwang@njupt.edu.cn, guiguan@njupt.edu.cn, hsari@ieee.org).

Yun Lin is with the College of Information and Communication Engineering, Harbin Engineering University, Harbin 150009, China (e-mail: linyun@hrbeu.edu.cn).

Tomoaki Ohtsuki is with the Department of Information and Computer Science, Keio University, Yokohama, Kanagawa 223-8522, Japan (e-mail: ohtsuki@keio.jp).

Bamidele Adebisi is with the Department of Engineering, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester M1 5GD, United Kingdom (e-mail: b.adebisi@mmu.ac.uk).

In this paper, a collaborative and cross-environment UAV classification method (termed FedUAV) based on federated semantic regularization (FedSR) is proposed. FL is used to obtain a UAV classification model with good classification performance and strong environment adaptability. In addition, semantic regularization is introduced into the FL process, making the learning process more stable and faster. The main contributions are summarized as follows:

- We proposed a novel method for classifying UAVs across various environmental conditions, employing the principles of federated learning to enhance both accuracy and privacy. The proposed approach allows multiple UAV surveillance systems to collaboratively learn a shared classification model without exchanging raw data. This is particularly important for operations requiring high confidentiality and data security.
- We proposed federated semantic regularization, where each participating surveillance system adjusts its local model by emphasizing the consistency between local features and global semantic prototypes. This regularization helps in mitigating the overfitting of models to local noise, generalizing well across more diverse environments. Our federated learning framework facilitates the aggregation of these semantically enriched local models into a robust global model.
- We conducted extensive experiments which demonstrate the efficacy of the method in consistently classifying UAVs under varying environmental conditions, significantly outperforming existing non-federated and federated models. These findings not only validate the robustness and adaptability of our approach but also illustrate its potential in surveillance applications where UAVs must operate effectively across diverse and challenging environments.

The remainder of this paper is organized as follows: Section II presents the related works. Section III describes the system model, the problem formulation, and the dataset. Section IV introduces our collaborative and cross-environment UAV classification approach. Section V presents the numerical results and discussions for a set of 6 UAVs in a cross-environment scenario. Finally, Section VI offers concluding remarks and outlines future research directions.

II. Related Works

A. RF-based UAV Detection and/or Classification

The paired controller and UAV form an entire communication system, and there are two way communication composed of uplink and downlink communication. By the uplink communication, the commands are sent to the UAV via controller; by the downlink communication, the UAV telemetry (flight data) and the video images from the payload are sent to the controller [11]. Both uplink signals and downlink signals can be utilized to detect and/or classify the UAVs.

Experimentally, it was observed that it is much more effective to detect the signal from the UAV controller as against the transmission from the UAV itself because the former has higher energy than the latter [12]. Therefore, there are some works [12], [13] that focus exclusively on the detection/classification of the RF signals from the controller. Ezuma et al. [12] split raw RF signals into frames and transformed them into the wavelet domain to remove the bias in signals and reduce the size of data to be processed; a naive Bayes approach is used to check for the presence of a UAV in each frame. In addition, a set of statistical features (i.e., skewness, variance, energy spectral entropy, and kurtosis) of energy transient signals is extracted and significant features are selected by performing neighborhood component analysis and subsequently are fed to several machine learning (ML) algorithms for classification. Ezuma et al. [13], [14] extended their work for identifying UAV controllers in the presence of wireless interference, i.e., Wi-Fi and Bluetooth devices. Not only is the scene more complex, but the number of statistical features is expanded from 4 to 15; the number of ML algorithms has been expanded from 4 to 5; the number of controllers has been expanded from 14 to 17; the confusion that results when attempting to classify UAV controllers of the same make and model is analyzed. For the same dataset [14], Bremnes et al. [15] decomposed RF signals into 16 fixed boundary empirical wavelet sub-band signals, and fed them into a lightweight deep convolutional neural network to classify various types of UAVs.

The passive UAV surveillance system listens for the RF signals in the monitored area. Affected by building occlusion, the surveillance system receives downlink video signals transmitted from the UAV with a higher probability than uplink control signals [16]. In addition, malicious users can conduct criminal activities through remote control to avoid detection, which means that the controller is far away from the surveillance system. This makes capturing the uplink signals for analysis potentially infeasible [17]. Therefore, there are some works that focus exclusively on the detection/classification of the RF signals from the UAV. Al-Sa'd et al. [18] captured a dataset [19] of 3 UAVs functioning in different modes, including off, on and connected, hovering, flying, and video recording. Subsequently, they computed the discrete Fourier transform (DFT) of each recorded segment, and then used three deep neural networks to detect the presence of a UAV, the presence of a UAV and its type, and lastly, the presence of a UAV, its type, and flight mode. After that, some neural networks that could further improve the classification performance on the dataset [19] were proposed, such as one dimensional convolutional neural network [20], multi-channel one-dimensional convolutional neural network [21], frequency-isolated multi-channel deep neural network [22], and time-frequency multiscale convolutional neural network [23]. Xue et al. [17] made a performance improvement of the UAV classification system from the aspects of signal preprocess, signal representation, signal augmentation and deep neural networks.

With more numerous UAVs, more complex electromagnetic environments, and more demanding surveillance requirements, some delicate works have been proposed. Multi-classifier [24], stacked denoising autoencoder [25], [26] and multi-scale convolutional architecture [27] have been presented for a more detailed UAV detection or classification. Model-agnostic meta-learning [28], tri-residual semantic network [29] have been used to deal with UAV classification in a scenario where limited training samples are available. The ISM bands are generally populated by several homogeneous and heterogeneous RF transmissions and the received RF signals are composed of multiple transmissions. According to this, YOLO-lite architecture that performs the UAV detection and classification on the spectrogram image and a spectrogram segmentation method that directly splits the entire spectrum into several subspectrograms to separate the interference signals working outside of the UAV bandwidth were presented in [30] and [16], respectively. For a noise-robust UAV classification, Chen et al. [31] proposed a threshold calculation algorithm based on global context information, reducing the influence of the noise on the UAV classification. For a reliable UAV classification, Chen et al. [32] proposed a generalized Pareto distribution model-based uncertainty scoring function, enabling an ability to detect both out-of-distribution and misclassified samples.

B. Federated Learning with RF Signal Processing

Most of current RF signal processing methods such as modulation classification [33]–[36], RF fingerprinting [37], jamming recognition [38], [39] introduced the FedAvg [9] to enable distributed schemes, and then paid attention to signal representation, lightweight network design and privacy protection. Wang et al. [33] proposed a distributed modulation classification scheme based on FedAvg and convolutional neural network and they extended an approach [34] based on balanced cross entropy for the condition of class imbalance and noisy varying. Liu et al. [35] designed a feature fusion network and signal representation composed of generalized envelope square spectrum and fractional lower order cyclic spectrum for a distributed modulation classification scheme. Liu et al. [37] designed a deformable convolutional network for distributed RF fingerprinting identification. Liu et al. [38] used a densely connected convolutional network and signal representation composed of Wegener-willie distribution, fractional Fourier transform and constellation diagram for a distributed few-shot jamming recognition. Meftah et al. [39] presented spectral correlation function and convolutional neural network-based distributed jamming recognition. For privacy protection, differential privacy federated learning [36] and blockchain [40], [41] have been introduced into the distributed modulation classification method.

These works have high communication overhead due to frequent communication between servers and clients, whereas these neural networks have high space complexity. Increasing communication interval [9] and network lightweight design [42]–[44] have been suggested as the solutions. Fu et al. [42] designed a lightweight network based on separable convolution for distributed modulation classification. Dong et al. [43] designed a lightweight network composed of a phase estimator and transformer, spatial feature extractor, and temporal feature extractor for distributed modulation classification. Shi et al. [44] modified a residual network for distributed RF fingerprinting.

In addition, these methods are at risk of declining classification performance in a scenario with data heterogeneity because FedAvg is quite challenging to deal with data heterogeneity, which introduces a drift in the updates of each client, resulting in slow and unstable convergence [45]. Data sharing and the dispersion regularization between global and local models have been suggested as the solutions. Reus-Muns et al. [10] leveraged the data-sharing approach in [46] to relax the data heterogeneity in distributed UAV classification. Qi et al. [47] designed a conditional variational autoencoder to generate a synthetic modulation dataset, which is stored in the server and available for each client with uploading a portion of its private data, mitigating the data heterogeneity significantly. Qi et al. [48] regarded the private classes of a particular local device as incremental classes and incremental learning was adopted to learn the classification knowledge of private classes, avoiding excessive dispersion between the global model parameters and local model parameters. Zhang et al. [49] used the proximal term [50] to ensure that the local model maintains a finite distance from the global model during learning.

C. Discussion and Motivation

Current UAV classification typically assumes that sufficient samples are available at a central location for training. Yet, the resulting models are generally only applicable within their training environments, making them difficult to apply to location-flexible UAVs. Federated learning provides an effective solution to obtain a generalized UAV classification model across environments by combining samples from multiple environments. However, few studies have focused on federated UAV classification schemes. Furthermore, to the best of our knowledge, no existing literature has analyzed feature drift in heterogeneous RF signals collected from different environments in UAV classification. We have conducted a detailed analysis of this phenomenon and proposed a solution: federated semantic regularization.

III. System Model and Problem Formulation

A. System Model

The system, composed of multiple UAV surveillance systems and a parameter server, is shown in Fig. 1. These systems are located in multiple geo-locations, and UAV RF signals are collected and stored on each system. The parameter server schedules each surveillance system to perform local learning and enables federated optimization. Due to the cooperation between the systems and the parameter server, each local dataset is fully utilized to obtain a global model that generalizes for UAV classification.

B. Problem Formulation

1) DL-based UAV classification method: In the DL-based UAV classification method, the goal is to produce a model $\mathcal{F}$ with a mapping function from RF signal space to category space. The collected dataset D is drawn from the distribution $\mathcal{P}(\boldsymbol{x}, y)$, where $\boldsymbol{x}$ and y denote the RF signals and corresponding label, respectively. This model is parameterized

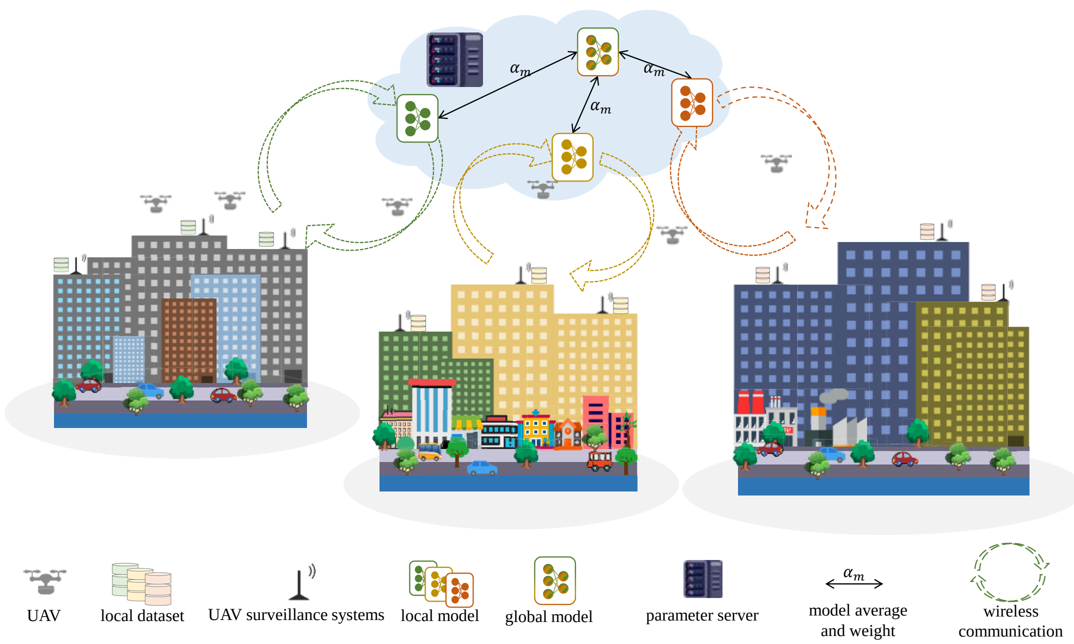


Fig. 1. The system model of proposed federated semantic regularization-based UAV classification method. The parameter server schedules each surveillance system to perform local learning and enables federated optimization. Due to the cooperation between the systems and the parameter server, each local dataset is fully utilized to obtain a global model that generalizes for UAV classification.

by learnable weights $\boldsymbol{w}$ and RF signals $\boldsymbol{x} \in D$. The objective function is

$$\arg\min_{\boldsymbol{w}} \mathcal{L}_{\boldsymbol{w}}(\mathcal{F}(\boldsymbol{w};\boldsymbol{x}), y), \tag{1}$$

where $\mathcal{L}_{\boldsymbol{w}}$ is a general definition of any supervised learning task (e.g., a cross-entropy loss).

2) FL-based UAV classification method: In the FL-based UAV classification method, each surveillance system owns a local dataset D_m drawn from the distribution $\mathcal{P}_m(\boldsymbol{x}, y)$. Usually, these systems share a model $\mathcal{F}$ with the same architecture and hyperparameters. This model is parameterized by learnable weights $\boldsymbol{\omega}$ and RF signals $\boldsymbol{x} \in D_m$. The objective function is

$$\arg\min_{\boldsymbol{w}} \sum_{m=0}^{M-1} \frac{|D_m| \left(\mathcal{L}_{\boldsymbol{w}}(\mathcal{F}(\boldsymbol{w};\boldsymbol{x}), y) + \lambda \mathcal{R}_{\boldsymbol{w}}(\mathcal{F}(\boldsymbol{w};\boldsymbol{x}), \mathcal{T})\right)}{\sum_{m=0}^{M-1} |D_m|}, \tag{2}$$

where M denotes the number of surveillance systems, $|D_m|$ is the number of RF signals in the local dataset, $\mathcal{R}_{\boldsymbol{w}}$ is a general definition of any regularization term, and λ represents the weight scalar to balance the two terms.

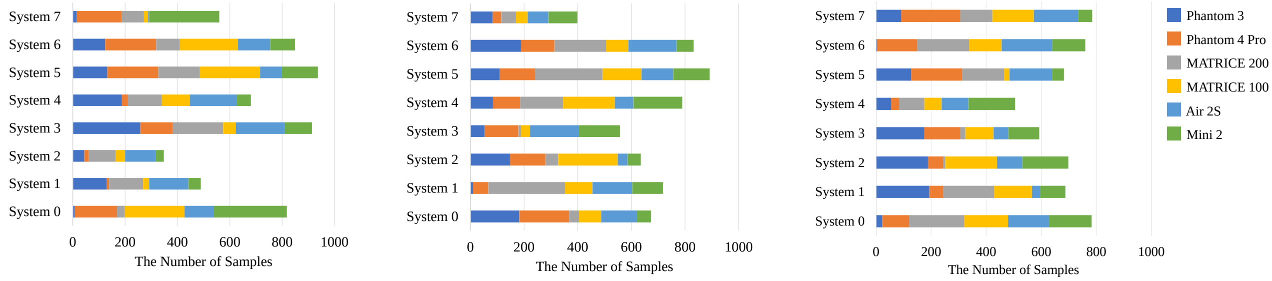
However, in a real-world FL environment, each system is deployed in a particular location, leading to a statistically heterogeneous environment. In the statistical heterogeneous settings, $\mathcal{P}_m$ varies across systems, indicating heterogeneous input/output space for $\boldsymbol{x}$ and y. Therefore, a federated semantic regularization-based UAV classification method is proposed, which can utilize these heterogeneous data to build a generalized model for UAV classification.

C. Dataset Description

DroneRFa [51] is a large-scale dataset of UAV RF signals. A software-defined radio device, USRP-2955, is used to monitor signals between UAVs and their controllers, including 9 types of flying UAV signals in an outdoor environment, 15 types of UAV signals in an indoor environment, and 1 type of background signal as reference. We select 6 types of flying UAV signals in an outdoor environment to build the distributed data scenario, that are Phantom 3, Phantom 4 Pro, MATRICE 200, MATRICE 100, Air 2S, and Mini 2. The RF signals of each UAV are collected when the distance between the UAV and its controller is 20-40 m (i.e, environment 0), 40-80 m (i.e., environment 1), and 80-150 m (i.e., environment 2), respectively. The sampling rate is 100 MS/s, the center frequency is 2,440 MHz, the receive gain is 50 dB. For each type of UAV at each distance, 8 signal segments with no less than 100 million sampling points are temporally collected. Therefore, we select the first 6 signal segments to build the distributed data scenario, using the remaining 2 segments as the testing dataset. These segments are split into multiple samples containing 1 million sampling points, and there is no overlap between any two samples. Finally, the details of distributed data scenario are shown in Fig. 2.

IV. FedSR-based UAV Classification Method

In the FedSR-based UAV classification method, the parameter server filters out a valid subset $\hat{M}$ from all the surveillance systems M as participants in this round of FL. After the participants are selected, all the systems in the participant subset will receive the model with global weights $\boldsymbol{w}$ from the server, and perform **local learning** (see the details in Section IV-A) driven by the local dataset to update the weights. ShuffleNetV2 [52], a lightweight neural network, is an example of the model. After obtaining the model with updated weights $\boldsymbol{w}_i$, the parameters are sent to the server



(a) Environment 0

(b) Environment 1

(c) Environment 2

Fig. 2. The sample distribution of local datasets of UAV surveillance systems. Specifically, the RF signals from UAVs, collected in various environments, are divided into 24 subsets to simulate 24 UAV surveillance systems and their local datasets.

through the communication network. The server performs **federated optimization** (see the details in Section IV-B), aggregating the received weights to obtain the global weights $\boldsymbol{w}$, then returns the global weights to each system for the next round of FL. A generalized global model is obtained after such rounds are performed sufficiently often.

The statistical heterogeneity of UAV RF signals shows feature drift in a low-dimensional manifold. Therefore, we design FedSR to alleviate feature drift. A semantic prototype of each UAV is learned from local learning on each system and aggregated from federated optimization on the server. Each system can improve the generalization ability of the local model through weight interactions and semantic prototype interactions. More details are described below.

A. Local Learning

1) RF signal preprocessing: The received RF signals are transformed by short time Fourier transform (STFT), which can be formulated as

$$X_{k,m} = \sum_{n=0}^{N-1} x[n]\omega[n-mR]e^{-j\frac{2\pi}{N}kn} \quad \text{for} \quad k=1,2,\dots,N \quad \text{and} \quad m=1,2,\dots,M, \tag{3}$$

where $\omega[n-mR]$ is a Hamming window function that splits the signal into multiple segments, mR is the center of Hamming window, and $\mathbf{X}$ denotes a two-dimensional complex matrix with each column representing the Fourier-transformed sequence of each signal segment, $X_{k,m}$ is the element of the matrix $\mathbf{X}$, M is number of columns of $\mathbf{X}$, N is the number of rows of $\mathbf{X}$. The width of Hamming window is 1,024, and the R is 1,024. The spectrum in dB scale, $\widetilde{\mathbf{X}}$ is given as

$$\widetilde{\mathbf{X}} = 10\log_{10}\left(|\mathbf{X}|^2\right), \tag{4}$$

where $|\mathbf{X}|$ is the amplitude of $\mathbf{X}$. The dimension of the spectrum is 1024×976. To facilitate the neural network with a larger batch size, the spectrum is downsampled to $1/4$ of the original spectrum and cut to 244×244. The final spectrum of RF signal of each UAV is shown in Fig. 3

2) Local optimizer: The classification performance of the current local model on the local dataset is measured by the cross-entropy loss, which can be formulated as

$$\mathcal{L}_{\boldsymbol{w}_m} = \frac{1}{|D_m|}\sum_{d=0}^{|D_m|-1} -\log\frac{\exp\left(q_{d,y_d}\right)}{\sum_{c=0}^{C-1}\exp\left(q_{d,c}\right)}, \tag{5}$$

where $q_{d,c}$ is an element of the vector $\mathbf{q_d}$ produced by $\mathcal{F}(\boldsymbol{w}_m^t;\widetilde{\boldsymbol{X}}_d)$, C is the number of categories of UAV. For each local dataset, the local semantic prototypes $\overline{\boldsymbol{P}}_m^t = \{\overline{\boldsymbol{p}}_{m,c}^t | c = 0,1,...,C-1\}$ can be calculated by

$$\boldsymbol{p}_{m,c}^t = \frac{\sum_{d=0}^{|D_m|-1}\mathbb{1}\left\{\mathrm{y}_d == c\right\}\boldsymbol{z}_d}{\sum_{d=0}^{|D_m|-1}\mathbb{1}\left\{\mathrm{y}_d == c\right\}}, \quad c=0,1,\dots,C-1, \tag{6}$$

where $\boldsymbol{z}_d$ is a feature vector produced by the submodule of $\mathcal{F}(\boldsymbol{w}_m^t;\widetilde{\boldsymbol{X}}_d)$. The feature drift of the RF signal across environments is measured by the mean square error between the feature vectors and the global semantic prototypes $\overline{\boldsymbol{P}}^t = \{\overline{\boldsymbol{p}}_c^t | c=0,1,...,C-1\}$, which can be formulated as

$$\mathcal{R}_{\boldsymbol{w}_m} = \frac{1}{|D_m|}\sum_{d=0}^{|D_m|-1}\left(\boldsymbol{z}_d - \overline{\boldsymbol{p}}_{y_d}^t\right)^2, \tag{7}$$

where $\overline{\boldsymbol{p}}_{y_d}^t$ is the global semantic prototype of category y_d and is updated during federated optimization. Therefore, we can obtain a local model with better classification performance and avoid feature drift by minimizing these two losses, which can be formulated as

$$\min \mathcal{L}_{\boldsymbol{w}_m} + \lambda\mathcal{R}_{\boldsymbol{w}_m}, \tag{8}$$

where λ represents the weight scalar to balance the two losses. The illustration of effectiveness of FedSR is shown in Fig. 4.

Stochastic gradient descent (SGD) with momentum is adopted to solve the optimization problem. The gradient of losses with respect to each weight is computed and the weights are modified along the downhill direction of the gradient in order to reduce the losses. That is, the modification of the weight vector at the current time step depends on both the current gradient and gradient change of the previous step as

$$\boldsymbol{g}_m^{t+1} \leftarrow \nabla_{\boldsymbol{w_m}}\left(\mathcal{L}_{\boldsymbol{w}_m} + \lambda\mathcal{R}_{\boldsymbol{w}_m}\right), \tag{9a}$$

$$\boldsymbol{v}_m^{t+1} \leftarrow \alpha\boldsymbol{v}_m^t - \epsilon\boldsymbol{g}_m^{t+1}, \tag{9b}$$

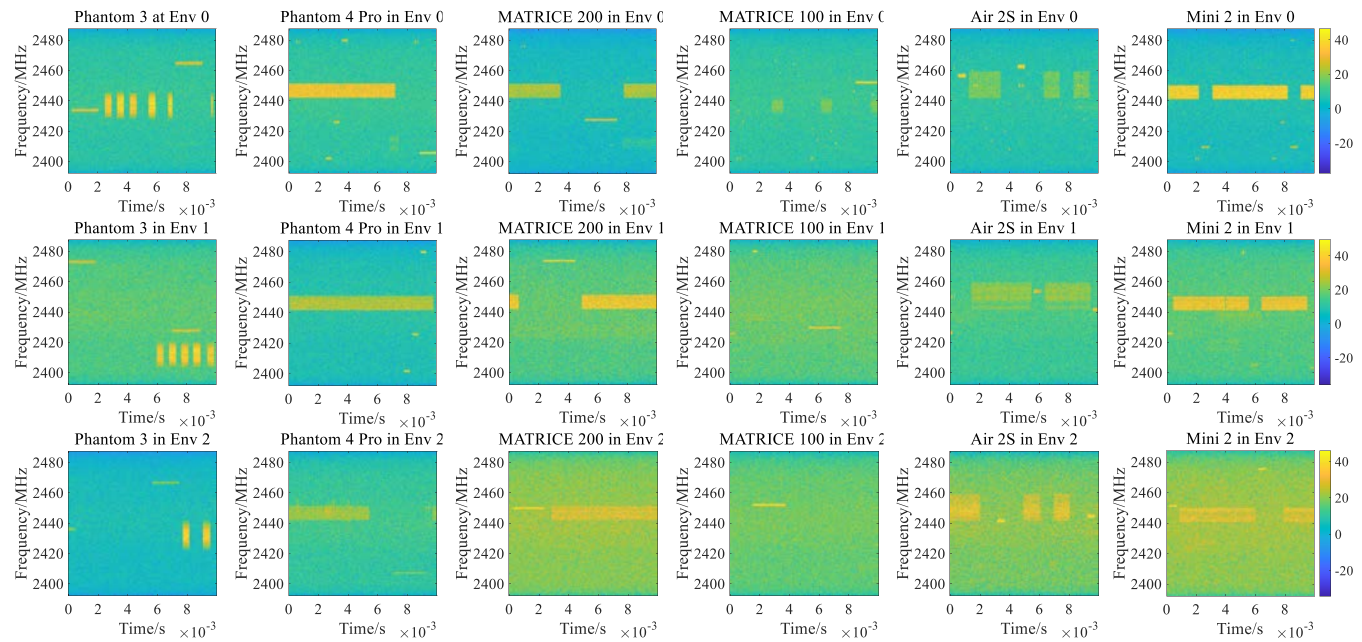


Fig. 3. The spectrum of RF signal from each UAV in various environments. Since the color bars of these spectrums are the same, only three color bars are showed for a better layout. Furthermore, the color bar indicates the amplitude of the spectrum, which is formulated in Equation (4).

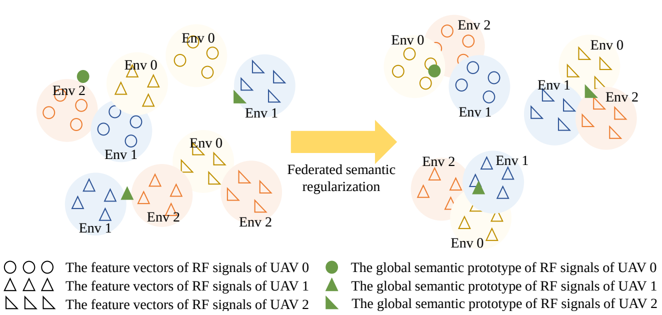


Fig. 4. The illustration of effectiveness of FedSR.

$$\boldsymbol{w}_m^{t+1} \leftarrow \boldsymbol{w}_m^{t} + \boldsymbol{v}_m^{t+1}, \tag{9c}$$

where α is the momentum parameter, ϵ is the learning rate. After iterating the local optimizer multiple times, the weights of the local model and the local semantic prototypes are uploaded to the server to perform a federated optimization.

B. Federated Optimization

Local learning establishes the mapping relationship between the features of the RF signal of the UAV, collected in different environments, and their categories. The mapping relationship is constructed by the weights of each local model. By aggregating the weights of each local model to get global weights, federated optimization has the potential to further construct the mapping relationship between the environment-independent features of the RF signal of the UAV and its categories. In addition, the federated optimization by aggregating the local semantic prototypes to get the global semantic prototypes, which provide a reference in the process of learning the features of the RF signal of the UAV in each local model, prompts the features of the RF signal of the same UAV collected in different environments close to the reference. The details of weight aggregation and prototype aggregation are as follows.

- Weight aggregation: The server takes the average of weights of the resulting local models, that is

$$\boldsymbol{w}^{t+1} \leftarrow \frac{1}{\hat{M}} \sum_{m=0}^{\hat{M}-1} \boldsymbol{w}_m^{t+1}. \tag{10}$$

- Semantic prototype aggregation: The local semantic prototypes belonging to various local dataset are averaged on the server to obtain the global semantic prototypes $\overline{\boldsymbol{P}}^{t+1} = \{\overline{\boldsymbol{p}}_c^{t+1} | c = 0, 1, ..., C-1\}$, that is

$$\overline{\boldsymbol{p}}_c^{t+1} \leftarrow \frac{1}{\hat{M}} \sum_{m=0}^{\hat{M}-1} \boldsymbol{p}_{m,c}^{t+1}, c = 0, 1, ..., C-1. \tag{11}$$

In **Algorithm 1**, we summarize the process of FedSR-based UAV classification method.

V. Numerical Result and Discussion

A. Simulation Parameters

We implemented our approach in PyTorch [53] (v1.10.2 with Python 3.6.13). The learning rate ϵ is 0.01 and the momentum parameter α is 0.9. The weight scalar λ is 0.2 and the aggregated weight scalar α_m is $\frac{1}{\hat{M}}$. The number of local epoch T_l is 5. The value of fraction of clients C_f is 0.1. We train the model for 300 communication rounds and the batch size is 16. Experiments are performed using NVIDIA GeForce RTX 3090 GPU.

Algorithm 1: FedSR-based UAV classification method

B is the local minibatch size; T_l is the number of local epochs; T is the number of communication round; C_f is the fraction of systems that perform local learning on each round.

[Server executes]

initialize $\boldsymbol{w}^0$, $\overline{\boldsymbol{P}}^0$

for *each round* $t = 0$ *to* $T-1$ **do**

  $\hat{M} \leftarrow \max(C_f \cdot M, 1)$

  $S_t \leftarrow$ random set of $\hat{M}$ systems

  for *each system* $m \in S_t$ *in parallel to* **do**

    $\boldsymbol{w}_m^{t+1}, \boldsymbol{P}_m^{t+1} \leftarrow \text{SystemUpdate}(m, \boldsymbol{w}^t, \overline{\boldsymbol{P}}^t)$

  end

  Weight aggregation as Equation (10)

  Semantic prototype aggregation as Equation (11)

end

SystemUpdate$(m, \boldsymbol{w}^t, \overline{\boldsymbol{P}}^t)$

$\boldsymbol{w}_m^t \leftarrow \boldsymbol{w}^t$

for *each local epoch* $t_l = 0$ *to* $T_l - 1$ **do**

  $\mathcal{B} \leftarrow$ Split D_m into batches of size B

  for *batch* $b \in \mathcal{B}$ **do**

    Calculate cross-entropy loss as Equation (5)

    Calculate feature drift as Equation (7)

    Update local model as Equation (9)

  end

end

Compute local semantic prototypes $\overline{\boldsymbol{P}}_m^{t+1}$ as Equation (6)

return $\boldsymbol{w}_m^{t+1}$ and $\overline{\boldsymbol{P}}_m^{t+1}$ to the server

B. Benchmarks

The proposed method is compared with multiple benchmarks. To ensure the fairness of the comparison, without special explanation, the model, loss function, optimizer, and learning rate used by each benchmark is ShuffleNetV2, cross-entropy loss, SGD with momentum, and 0.01, respectively.

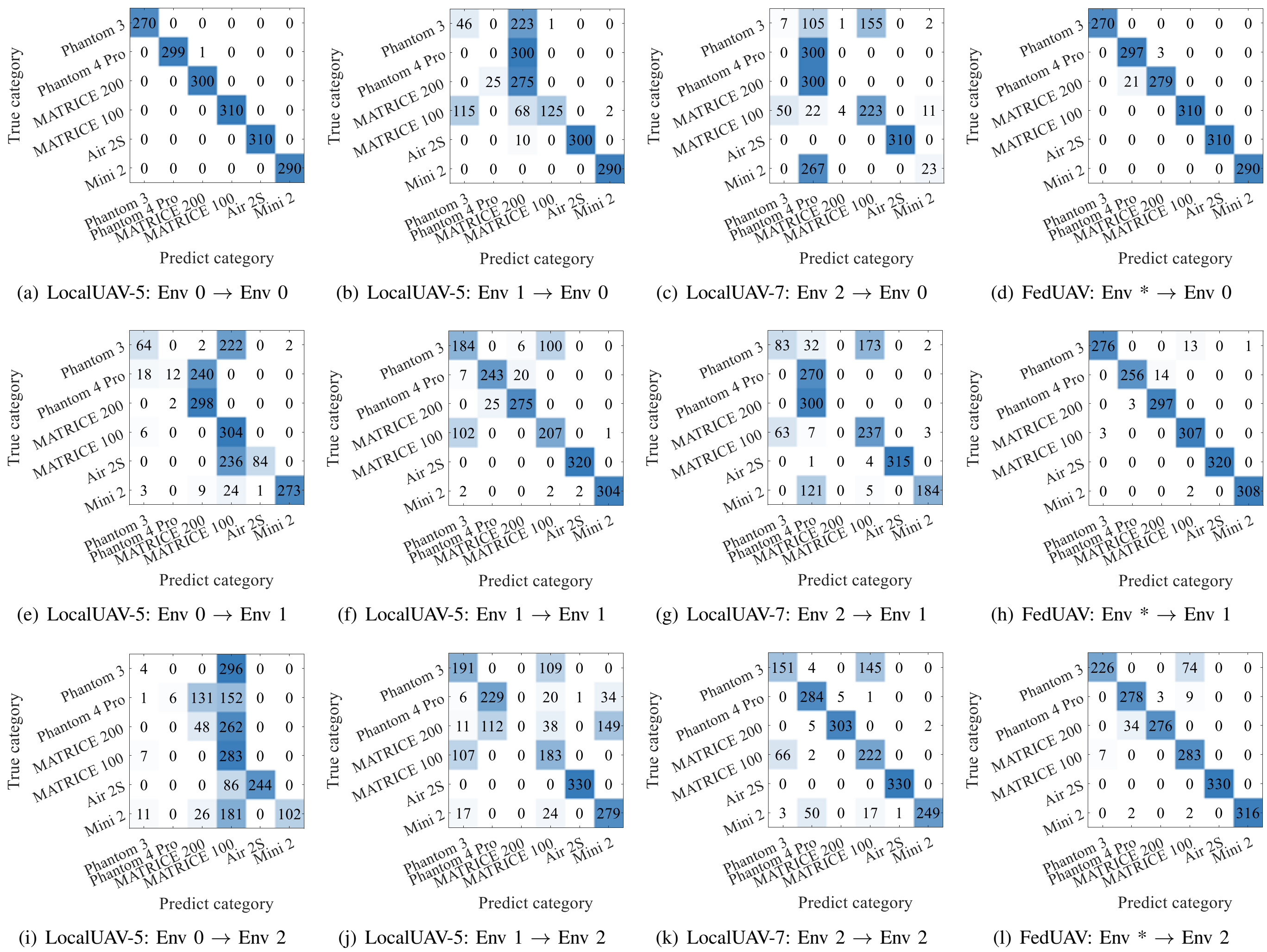
- CentUAV: Multiple suveillance systems transmit their local dataset to the server. The server uses the integrated dataset to perform the training process for obtaining an UAV classification model.
- LocalUAV: The surveillance system uses its own local dataset to perform the training process for obtaining an UAV classification model.
- LocalUAV+ (ResNet18) [24], [51]: The surveillance system uses its own local dataset to perform the training process for obtaining an UAV classification model. Specifically, the model used here is ResNet18, which is more complex than ShuffleNetV2.
- LocalUAV+ (Triplet Loss) [29]: The surveillance system uses its own local dataset to perform the training process for obtaining an UAV classification model. Specifically, the loss function used here is the combination of cross-entropy loss and triplet loss.
- FedAvg [9]: Multiple surveillance systems perform local training and the server performs federated optimization. FedAvg differs from FedSR in that the local training of FedAvg is supervised by cross-entropy loss, and federated optimization of FedAvg consists of weight aggregation.
- FedSGD [33]: Multiple surveillance systems perform local training and the server performs federated optimization. FedSGD differs from FedSR in that all surveillance systems of FedSGD participate in local training and federated optimization at each communication round, the local training is supervised by cross-entropy loss, and federated optimization consists of weight aggregation.
- FedProx [49]: Multiple surveillance systems perform local training and the server performs federated optimization. FedProx differs from FedSR in that the local training of FedProx is supervised by cross-entropy loss and proximal term, and federated optimization of FedProx consists of weight aggregation.

C. Classification Accuracy (ACC)

1) FedUAV vs. LocalUAV: The ACCs across three different environments are shown in Table I. LocalUAVs exhibit varied ACC across the three environments, with each LocalUAV achieving its highest ACC within the environment it was presumably optimized for. However, there is a noticeable drop in ACC when these LocalUAVs are tested in environments other than their primary one, indicating a specialization that comes at cost of generalization. Across all the environments, the ACC of FedUAV significantly surpasses most of LocalUAVs, with ACC exceeding 92%. This suggests that FedUAV has superior generalization capabilities and adaptability to different environments.

The confusion matrices of LocalUAV-5 from environment 0, LocalUAV-5 from environment 1, LocalUAV-7 from environment 2, and FedUAV are shown in Fig. 5, where these LocalUAVs demonstrate superior performance in their respective environments, compared to the other LocalUAVs in the same environments. We also observed that there is a noticeable drop in performance when these LocalUAVs are tested in environments other than their designated ones. The FedUAV can correctly classify the UAVs across all the environments. Notably, the ACC of confusion matrix of FedUAV is inconsistent with that in Table I because the ACC in Table I is the average ACC of 5 experiments, while the confusion matrix is one of the 5 experiments.

2) FedUAV vs. Existing UAV classification methods: Some advanced neural networks and loss functions have been employed to enhance UAV classification (termed LocalUAV+ in this paper), such as the ResNet18 [24], [51] and triplet loss [29]. As shown in TABLE II, LocalUAV+ achieves higher ACC than LocalUAV. However, ACC still drop significantly when LocalUAV+ is tested in environments other than it designated ones. The superiority of FedUAV over LocalUAV has been shown in TABLE I, and TABLE II further demonstrates its advantages over the existing FL algorithms, such as FedAvg [9], FedSGD [33] and FedProx [49]. FedSR shows better robustness to environments. CentUAV has the highest ACC in this scenario. It achieves superior classification performance with substantial communication overhead and potential privacy risks. The proposed FedSR-enabled FedUAV approaches



(a) LocalUAV-5: Env 0 $\rightarrow$ Env 0 (b) LocalUAV-5: Env 1 $\rightarrow$ Env 0 (c) LocalUAV-7: Env 2 $\rightarrow$ Env 0 (d) FedUAV: Env * $\rightarrow$ Env 0

(e) LocalUAV-5: Env 0 $\rightarrow$ Env 1 (f) LocalUAV-5: Env 1 $\rightarrow$ Env 1 (g) LocalUAV-7: Env 2 $\rightarrow$ Env 1 (h) FedUAV: Env * $\rightarrow$ Env 1

(i) LocalUAV-5: Env 0 $\rightarrow$ Env 2 (j) LocalUAV-5: Env 1 $\rightarrow$ Env 2 (k) LocalUAV-7: Env 2 $\rightarrow$ Env 2 (l) FedUAV: Env * $\rightarrow$ Env 2

Fig. 5. The confusion matrices of LocalUAV-5 from environment 0, LocalUAV-5 from environment 1, LocalUAV-7 from environment 2, and FedUAV. "Env $number_1 \rightarrow$ Env $number_2$" means that the UAV classification model is trained using the dataset collected from the environment $number_1$ but tested using the dataset collected from the environment $number_2$. "Env * $\rightarrow$ Env $number$" means that the UAV classification model is trained using the distributed dataset but tested using the dataset collected from the environment $number$.

CentUAV's performance, offering a viable alternative with fewer privacy risks and communication overhead.

3) FedSR vs. FedAvg: We further analyzed the ACC of FedAvg and FedSR across various regularization intensities, represented by the value of λ in Equation (8). As demonstrated in Fig. 6, FedSR consistently outperforms FedAvg across many intensities. Proper regularization effectively guides the local training of FedSR, enhancing the classification performance of local models on local datasets without significant deviation from the global model. This improves the model's robustness across different environments. However, the local training of FedAvg puts more emphasis on the classification performance of local models on local datasets, so it deviates from the global model, resulting in the model being less robust to the environment than FedSR. When the proposed FedSR is compared to the benchmarks, the λ used in FedSR is 0.2.

In addition to the performance advantages, FedSR has a more stable and faster convergence process than FedAvg. The loss convergences of several difficult systems are shown in Fig. 7, that are system 2 in environment 0, system 3 in environment 1, and system 4 in environment 2. These systems have limited samples, and the number of samples of each category varies greatly, showing a long-tail distribution.

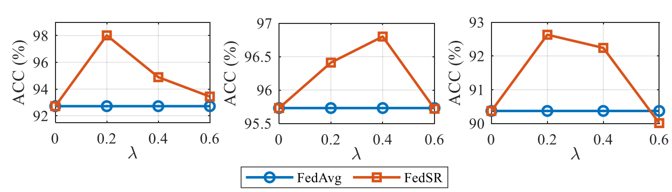


Fig. 6. The ACC of proposed FedSR and FedAvg. The left, middle and right figures show the ACC of the models in environment 0, 1 and 2, respectively.

D. Effectiveness Analysis of Semantic Regularization

We attempted to explain the misclassification of LocalUAVs and the classification of FedUAV by analyzing the feature drift of testing samples. We also chose LocalUAV-5 from environment 0, LocalUAV-5 from environment 1, and

TABLE I
THE ACC OF FEDUAV AND LOCALUAV

		Env 0	Env 1	Env 2
Environment 0	LocalUAV-0	82.42	53.61	32.34
	LocalUAV-1	75.62	38.56	38.64
	LocalUAV-2	90.67	41.83	34.84
	LocalUAV-3	95.28	68.83	30.11
	LocalUAV-4	97.92	51.00	31.20
	LocalUAV-5	**99.94**	57.50	37.34
	LocalUAV-6	99.55	46.33	30.60
	LocalUAV-7	99.21	58.83	39.67
	LocalUAV (mean)	92.58	52.06	34.34
Environment 1	LocalUAV-0	66.01	78.89	56.63
	LocalUAV-1	65.96	78.50	57.39
	LocalUAV-2	67.64	73.83	53.04
	LocalUAV-3	73.88	67.56	63.26
	LocalUAV-4	66.18	84.11	59.24
	LocalUAV-5	58.20	85.17	65.87
	LocalUAV-6	43.99	84.17	53.21
	LocalUAV-7	40.45	81.22	39.67
	LocalUAV (mean)	60.29	79.18	56.04
Environment 2	LocalUAV-0	76.57	67.44	72.66
	LocalUAV-1	61.52	65.83	76.52
	LocalUAV-2	75.34	54.61	75.71
	LocalUAV-3	59.38	65.33	73.97
	LocalUAV-4	62.42	54.56	67.23
	LocalUAV-5	64.78	47.17	72.93
	LocalUAV-6	60.56	66.17	78.48
	LocalUAV-7	48.48	60.50	83.64
	LocalUAV (mean)	63.63	60.20	75.14
/	FedUAV (Proposed)	98.02	**96.41**	**92.63**

Note: **Number** represents the top-1 ACC; number represents the top-2 ACC.

TABLE II
THE ACC OF FEDUAV AND EXISTING UAV CLASSIFICATION METHODS

		Env 0	Env 1	Env 2
LocalUAV		92.58	79.18	75.14
LocalUAV+	ResNet18 [24], [51]	97.51	85.86	81.98
	Triplet Loss [29]	97.54	85.95	81.87
FedUAV	FedAvg [9]	92.72	95.73	90.37
	FedSGD [33]	92.10	95.37	88.30
	FedProx [49]	93.50	95.61	91.92
	FedSR (Proposed)	98.02	96.41	92.63
CentUAV		**99.94**	**99.00**	**97.72**

Note: **Number** represents the top-1 ACC; number represents the top-2 ACC.

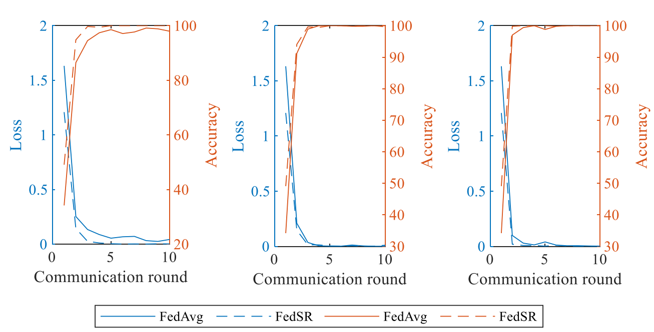


Fig. 7. The loss convergences of local models in the FedUAV. The left, middle and right figures show the loss convergences of the local models of system 2 in environment 0, the local models of system 3 in environment 1, and the local models of system 4 in environment 2, respectively.

LocalUAV-7 from environment 2 as representatives. These optimal local models and the optimal global model conduct tests on the RF signals from UAVs located across three different environments. These RF signals are transformed into features by the local models or global models, and subsequently reduced using T-distributed stochastic neighbor embedding (t-SNE) [54], which are then visualized in Fig. 8. The figures in the first row are colored according to the UAV from which the signal originates, the figures in the second row are colored according to the environment from which the signal originates, and the figures in the third row show the statistical results of the drift between these features and the corresponding prototype.

As shown in Fig. 8(a) to Fig. 8(c) and Fig. 8(e) to Fig. 8(g), for the same UAV signal encountered in different environments, the LocalUAV fails to extract consistent features. This indicates that the features of the same UAV signal in varying environments do not cluster together and instead overlap significantly with those of other UAV signals in the feature space. Notably, in addition to the visualization, we also quantified this inconsistency as shown in Fig. 8(i) to Fig. 8(k). Specifically, from the statistical analysis of the MSE between the features and its corresponding prototype, it can be seen that the kernel density curves of LocalUAV are flat, that is, the drift between the features and the prototype is large.

Conversely, as shown in Fig. 8(d) and Fig. 8(h), the FedUAV is capable of extracting consistent features, demonstrating that the features of the same UAV signal in various environments tend to converge and infrequently overlap with those of other UAVs in the feature space. Notably, we circled these consistent features with dashed lines in Fig. 8(h). From Fig. 8(l), it can be seen that the kernel density curves of FedUAV are steep, that is, the drift between the features and the prototype is small.

These visualization and statistical analyses effectively elucidate the classification performance of LocalUAV and FedUAV, and also confirm the presence of feature drift in LocalUAV, while the proposed method effectively mitigates this drift and enhances the generalization capabilities of the classification model.

E. Explanation Analysis of Classification Performance Across Environments

Further, we explored which areas of the input are emphasized by the FedUAV model during classification process and whether this focus remains consistent across different environments. We employed class activation mapping, as introduced by Zhou et al. [55], to visualize the regions of RF signals that the neural network targets to distinguish between different UAV types. As depicted in Fig. 9, FedUAV primarily targets the jump frequency blocks of RF signals for classification. This visualization confirms that FedUAV consistently focuses on these specific signal blocks when analyzing RF signals from the same UAV, regardless of the environmental context.

VI. CONCLUSION

This paper proposes a federated semantic regularization-based UAV classification method, showcasing significant

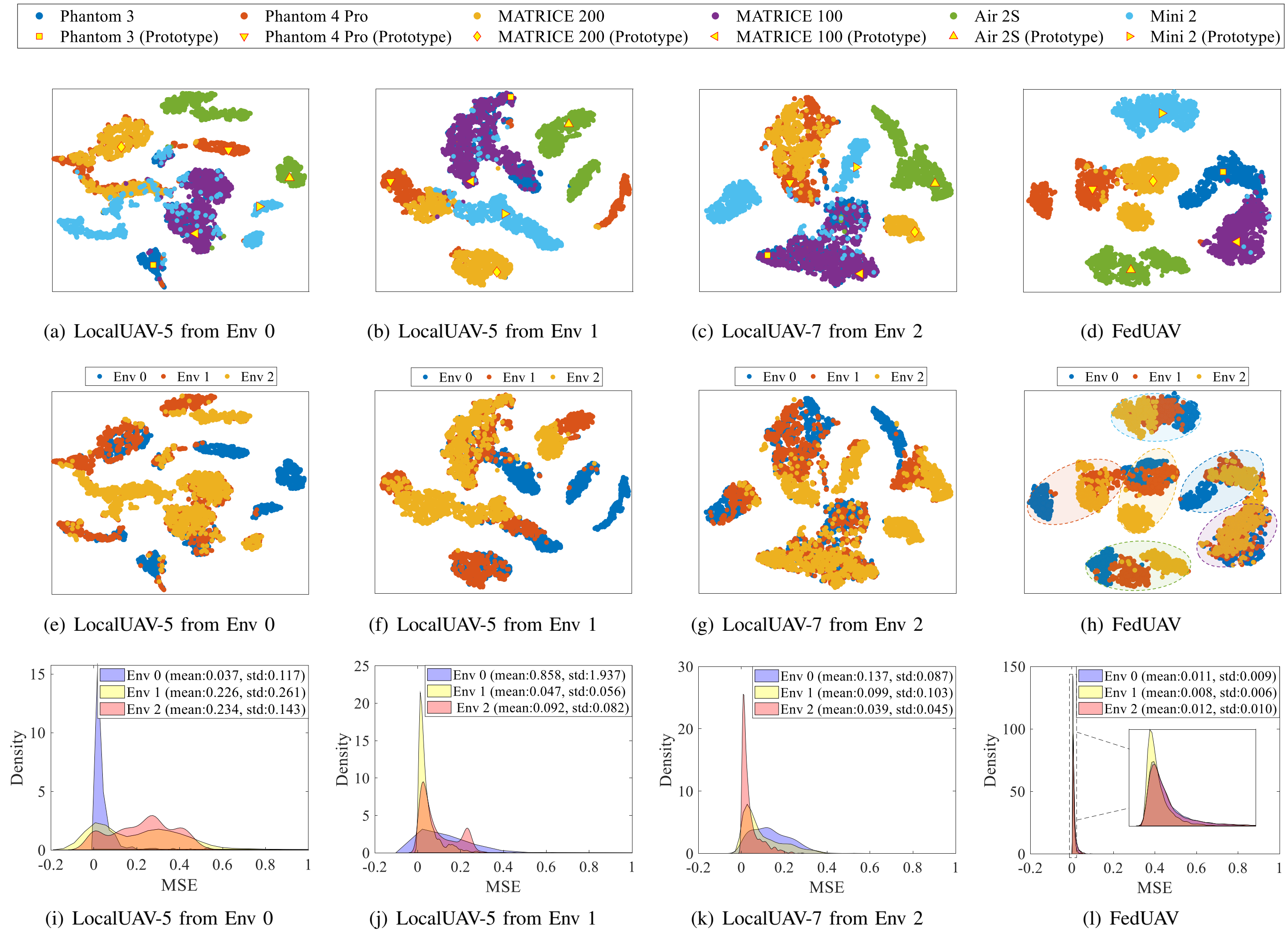


Fig. 8. The feature of the testing dataset across various environments, as extracted by LocalUAV and FedUAV. (c), (f), (i) and (l) are the statistical results of the drift between the features and the corresponding prototypes, where std is an abbreviation for the standard deviation.

improvements in model generalization across diverse environments. By emphasizing semantic consistency and utilizing a federated learning framework, the proposed method not only maintains high classification accuracy but also adapts seamlessly to new and changing environments. In addition, we carefully analyzed the feature drift in models that are trained in a specific environment but tested across environments. By using feature visualization and class activation mapping, we elucidated the mechanisms behind the generalization of the proposed method across different environments. In the future, models with the ability to generalize to unknown environments, which are unseen during the training process, will be explored.

References

[1] A. Fotouhi *et al.*, "Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3417–3442, Fourthquarter 2019.

[2] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated learning for internet of things: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1622–1658, Thirdquarter 2021.

[3] Z. Shi, M. Huang, C. Zhao, L. Huang, X. Du, and Y. Zhao, "Detection of LSSUAV using hash fingerprint based SVDD," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, 2017, pp. 1–5.

[4] C. Zhao, C. Chen, Z. Cai, M. Shi, X. Du, and M. Guizani, "Classification of small UAVs based on auxiliary classifier wasserstein GANs," in *Proc. IEEE Glob. Commun. Conf.*, Abu Dhabi, United Arab Emirates, 2018, pp. 206–212.

[5] P. Chen, *et al.*, "Point-to-box network for accurate object detection via single point supervision," in *Proc. Eur. Conf. Comput. Vis.*, Tel Aviv, Israel, October 23¨C27, 2022, pp. 51–67.

[6] Z. Wang, *et al.*, "Learning to detect head movement in unconstrained remote gaze estimation in the wild," in *Proc. IEEE Winter Conf. Appl. of Comput. Vis.*, Snowmass, CO, USA, 2020, pp. 3432–3441.

[7] X. Tu, *et al.*, "Image-to-video generation via 3D facial dynamics," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 32, no. 4, pp. 1805–1819, Apr. 2022.

[8] E. T. M. Beltrán, *et al.*, "Decentralized federated learning: Fundamentals, state of the art, frameworks, trends and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 2983–3013, Fourthquarter 2023.

[9] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. Statist.*, Fort Lauderdale, Florida, USA, 2017, pp. 1273–1282.

[10] G. Reus-Muns, and K. R. Chowdhury, "Classifying UAVs with proprietary waveforms via preamble feature extraction and federated learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6279–6290, Jul. 2021.

[11] İ. Çuhadar, and M. Dursun, "Unmanned air vehicle systems data links," *J. Autom. Control Eng.*, pp. 189–193, 2016.

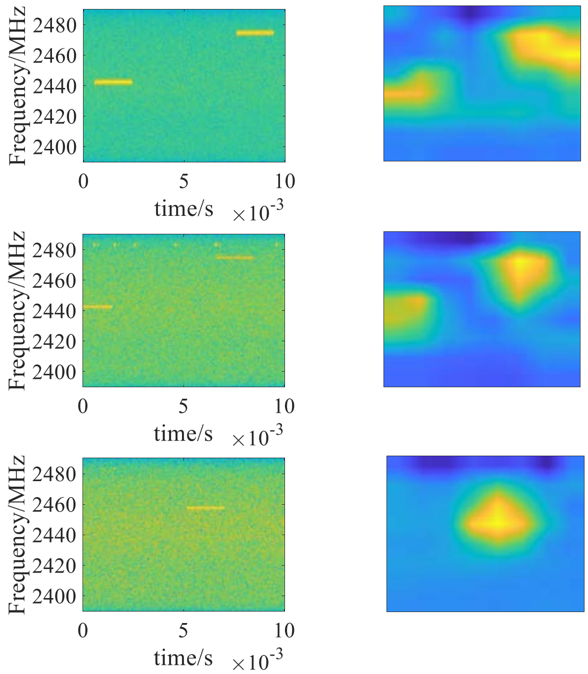


Fig. 9. The class activation mapping of the RF signal from the Phantom 3 UAV across various environments. The figures in the first line are the spectrum of RF signals of UAV located at environment 0 and its corresponding class activation mapping, respectively. The figures in the second line are the spectrum of RF signals of UAV located at environment 1 and its corresponding class activation mapping, respectively. The figures in the third line are the spectrum of RF signals of UAV located at environment 2 and its corresponding class activation mapping, respectively.

[12] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, and I. Guvenc, "Micro-UAV detection and classification from RF fingerprints using machine learning techniques," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, 2019, pp. 1–13.

[13] M. Ezuma, F. Erden, C. Kumar Anjinappa, O. Ozdemir, and I. Guvenc, "Detection and classification of UAVs using RF fingerprints in the presence of Wi-Fi and bluetooth tnterference," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 60–76, Nov. 2020.

[14] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, and I. Guvenc, "Drone remote controller RF signal dataset", *IEEE Dataport*, Nov. 2020.

[15] K. Bremnes, R. Moen, S. R. Yeduri, R. R. Yakkati, and L. R. Cenkeramaddi, "Classification of UAVs utilizing fixed boundary empirical wavelet sub-bands of RF fingerprints and deep convolutional neural network," *IEEE Sens. J.*, vol. 22, no. 21, pp. 21248–21256, Nov. 2022.

[16] H. Zhang, T. Li, Y. Li, J. Li, O. A. Dobre, and Z. Wen, "RF-based drone classification under complex electromagnetic environments using deep learning," *IEEE Sens. J.*, vol. 23, no. 6, pp. 6099–6108, Mar. 2023.

[17] C. Xue, T. Li, Y. Li, Y. Ruan, R. Zhang, and O. A. Dobre, "Radio-frequency identification for drones with nonstandard waveforms using deep learning," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, Aug. 2023.

[18] M. F. Al-Sa'd, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, "RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database," *Future Gener. Comput. Syst.*, vol. 100, pp. 86–97, May 2019.

[19] M. F. Al-Sa'd, *et al.*, "DroneRF dataset: A dataset of drones for RF-based detection, classification, and identification", *Mendeley Data*, Mar. 2019.

[20] S. Al-Emadi, and F. Al-Senaid, "Drone detection approach based on radio-frequency using convolutional neural network," in *Proc. IEEE Int. Conf. Informatics, IoT, Enabling Technol.*, Doha, Qatar, 2020, pp. 29–34.

[21] M. S. Allahham, T. Khattab, and A. Mohamed, "Deep learning for RF-based drone detection and identification: A multi-channel 1-D convolutional neural networks approach," in *Proc. IEEE Int. Conf. Informatics, IoT, Enabling Technol.*, Doha, Qatar, 2020, pp. 112–117.

[22] S. Yang, Y. Luo, W. Miao, C. Ge, W. Sun, and C. Luo, "RF signal-based UAV detection and mode classification: A joint feature engineering generator and multi-channel deep neural network approach," *Entropy*, vol. 23, no. 12, pp. 1678, 2021.

[23] S. Mandal, and U. Satija, "TimeÂ¨Cfrequency multiscale convolutional neural network for RF-based drone detection and identification," *IEEE Sens. Lett.*, vol. 7, no. 7, pp. 1–4, Jul. 2023.

[24] N. Soltani, G. Reus-Muns, B. Salehi, J. Dy, S. Ioannidis, and K. Chowdhury, "RF fingerprinting unmanned aerial vehicles with non-standard transmitter waveforms," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15518–15531, Dec. 2020.

[25] O. O. Medaiyese, M. Ezuma, A. P. Lauf, and A. A. Adeniran, "Cardinal RF (CardRF): An outdoor UAV/UAS/drone RF signals with bluetooth and WiFi signals dataset", *IEEE Dataport*, Jul. 2022.

[26] O. O. Medaiyese, M. Ezuma, A. P. Lauf, and A. A. Adeniran, "Hierarchical learning framework for UAV detection and identification," *IEEE J. Radio Freq. Identif.*, vol. 6, pp. 176–188, 2022.

[27] S. S. Alam, *et al.*, "RF-enabled deep-learning-assisted drone detection and identification: An end-to-end approach", *Sensors*, vol. 23, no. 9, pp. 4202, 2023.

[28] N. Yang et al., "Specific emitter identification with limited samples: A model-agnostic meta-learning approach," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 345–349, Feb. 2022.

[29] H. Liang, R. Wang, M. Xu, F. Zhou, Q. Wu, and O. A. Dobre, "Few-shot learning UAV recognition methods based on the tri-residual semantic network," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2072–2076, Sept. 2022.

[30] S. Basak, S. Rajendran, S. Pollin, and B. Scheers, "Combined RF-based drone detection and classification," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 1, pp. 111–120, Mar. 2022.

[31] Y. Chen, L. Zhu, C. Yao, G. Gui, and L. Yu, "Global context-based threshold strategy for drone identification under the low SNR condition," *IEEE Internet Things J.*, vol. 10, no. 2, pp. 1332–1346, Jan. 2023.

[32] Y. Chen, L. Zhu, Y. Jiao, C. Yao, K. Cheng, and Y. Gu, "An extreme value theory-based approach for reliable drone RF signal identification," *IEEE Trans. Cogn. Commun. Netw.*, vol. 10, no. 2, pp. 454–469, Apr. 2024.

[33] Y. Wang *et al.*, "Distributed learning for automatic modulation classification in edge devices," *IEEE Wireless Commun. Lett.*, vol. 9, no. 12, pp. 2177–2181, Dec. 2020.

[34] Y. Wang, G. Gui, H. Gacanin, B. Adebisi, H. Sari, and F. Adachi, "Federated learning for automatic modulation classification under class imbalance and varying noise condition," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 1, pp. 86–96, Mar. 2022.

[35] M. Liu, K. Yang, N. Zhao, Y. Chen, H. Song, and F. Gong, "Intelligent signal classification in industrial distributed wireless sensor networks based industrial internet of things," *IEEE Trans. Ind. Inform.*, vol. 17, no. 7, pp. 4946–4956, Jul. 2021.

[36] J. Shi, L. Qi, K. Li, and Y. Lin, "Signal modulation recognition method based on differential privacy federated learning," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–13, Oct. 2021.

[37] M. Liu, C. Liu, Y. Chen, Z. Yan, and N. Zhao, "Radio frequency fingerprint collaborative intelligent blind identification for green radios," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 940–949, Jun. 2023.

[38] M. Liu, Z. Liu, W. Lu, Y. Chen, X. Gao, and N. Zhao, "Distributed few-shot learning for intelligent recognition of communication jamming, " *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 3, pp. 395–405, Apr. 2022.

[39] A. Meftah, T. N. Do, G. Kaddoum, C. Talhi, and S. Singh, "Federated learning-enabled jamming detection and waveform classification for distributed tactical wireless networks," *IEEE Trans. Netw. Serv. Manag.*, vol. 20, no. 4, pp. 5053–5072, Dec. 2023.

[40] M. Umer, and C. S. Hong, "Blockchain-assisted ensemble federated learning for automatic modulation classification in wireless networks," in *Proc. KIISE Korea Comput. Congr.*, 2020, pp. 756–758.

[41] Z. Liu, J. Mu, W. Lv, Z. Jing, Q. Zhou, and X. Jing, "A distributed attack-resistant trust model for automatic modulation classification," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 145–149, Jan. 2023.

[42] X. Fu *et al.*, "Lightweight automatic modulation classification based on decentralized learning," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 1, pp. 57–70, Mar. 2022.

[43] B. Dong, *et al.*, "A lightweight decentralized-learning-based automatic modulation classification method for resource-constrained edge devices," *IEEE Internet Things J.*, vol. 9, no. 24, pp. 24708–24720, Dec. 2022.

[44] J. Shi, H. Zhang, S. Wang, B. Ge, S. Mao, and Y. Lin, "FedRFID: Federated learning for radio frequency fingerprint identification of WiFi

signals," in *Proc. IEEE Glob. Commun. Conf.*, Rio de Janeiro, Brazil, 2022, pp. 154–159.

[45] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 5132–5143.

[46] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," 2018, arXiv:1806.00582.

[47] P. Qi, X. Zhou, Y. Ding, Z. Zhang, S. Zheng, and Z. Li, "FedBKD: Heterogenous federated learning via bidirectional knowledge distillation for modulation classification in IoT-edge system," *IEEE J. Sel. Top. Signal Process.*, vol. 17, no. 1, pp. 189–204, Jan. 2023.

[48] P. Qi, X. Zhou, Y. Ding, S. Zheng, T. Jiang, and Z. Li, "Collaborative and incremental learning for modulation classification with heterogeneous local dataset in cognitive IoT," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 881–893, Jun. 2023.

[49] T. Zhang, D. Xu, P. Ren, K. Yu, and M. Guizani, "DFLNet: Deep federated learning network with privacy preserving for vehicular LoRa nodes fingerprinting," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2901–2905, Feb. 2024.

[50] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst. (MLSys)*, 2020, pp. 429–450.

[51] N. Yu, S. Mao, C. Zhou, G. Sun, Z. Shi, and J. Chen, "DroneRFa: A large-scale dataset of drone radio frequency signals for detecting low-altitude drones," *J. Electron. Inf. Technol.*, vol. 45, no. 11, pp. 1–9, 2023.

[52] N. Ma, X. Zhang, H-T Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, 2018, pp. 116–131.

[53] A. Paszke *et al.*, "Automatic differentiation in PyTorch," in *Proc. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 2017, pp. 1–4.

[54] L. Maaten, and G. Hinton,"Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 257–2605, Nov. 2008.

[55] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 2921–2929,