

Open Set Learning for RF-Based Drone Recognition via Signal Semantics

Ningning Yu^{ID}, *Student Member, IEEE*, Jiajun Wu^{ID}, Chengwei Zhou^{ID}, *Senior Member, IEEE*, Zhiguo Shi^{ID}, *Senior Member, IEEE*, and Jiming Chen^{ID}, *Fellow, IEEE*

Abstract—The abuse of drones has raised critical concerns about public security and personal privacy, bringing an urgent requirement for drone recognition. Existing radio frequency (RF)-based recognition methods follow the assumption of the closed set, resulting in the unknown signals being misclassified as known classes. To address this problem, we propose a Signal Semantic-based open Set Recognition (S3R) method in this paper. First, the short-time Fourier transform is introduced to construct the signal spectra, decoupling the drone signals with other interference signals. Then, we design a texture extractor and a position extractor to extract the texture features and position features from the spectra, respectively. The extracted features are further fused and structurally optimized to construct distinguishable signal semantics. Based on the structural characteristics of signal semantics, an outlier analysis-based semantic classifier is proposed, which searches the outliers of each known class in the closed set as the bounding thresholds to detect unknown instances. Finally, the detected unknown instances are further classified into their exact classes by implementing clustering in a new semantic space, where semantics are augmented by introducing basic features from the intermediate layers of the texture extractor. Besides, a real-world spectrogram dataset of commonly-used drones is released, which includes 24 classes and covers 7 brands. Extensive experiments demonstrate that the proposed S3R method outperforms the state-of-the-art methods in terms of accuracy and generalizability for both the closed set and the open set.

Index Terms—Drone recognition, open set recognition, RF signal recognition, signal semantic classification.

I. INTRODUCTION

IN RECENT years, drones, also known as unmanned aerial vehicles (UAVs), have shown significant growth in the commercial market and civilian popularity [1], [2],

Received 7 December 2023; revised 10 June 2024 and 19 July 2024; accepted 1 September 2024. Date of publication 20 September 2024; date of current version 29 October 2024. This work was supported in part by the National Natural Science Foundation of China under Grant U21A20456, in part by the Fundamental Research Funds for Central Universities under Grant 226-2023-00111 and Grant 226-2024-00004, and in part by Zhejiang University Education Foundation Qizhen Scholar Foundation. The associate editor coordinating the review of this article and approving it for publication was Prof. Danda Rawat. (*Corresponding author: Zhiguo Shi.*)

Ningning Yu, Jiajun Wu, Chengwei Zhou, and Zhiguo Shi are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China, and also with Jinhua Institute, Zhejiang University, Jinhua 321037, China (e-mail: nnyu@zju.edu.cn; wujiajun@zju.edu.cn; zhouchw@zju.edu.cn; shizg@zju.edu.cn).

Jiming Chen is with the State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China, and also with the School of Artificial Intelligence, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: jmchen@ieee.org).

Digital Object Identifier 10.1109/TIFS.2024.3463535

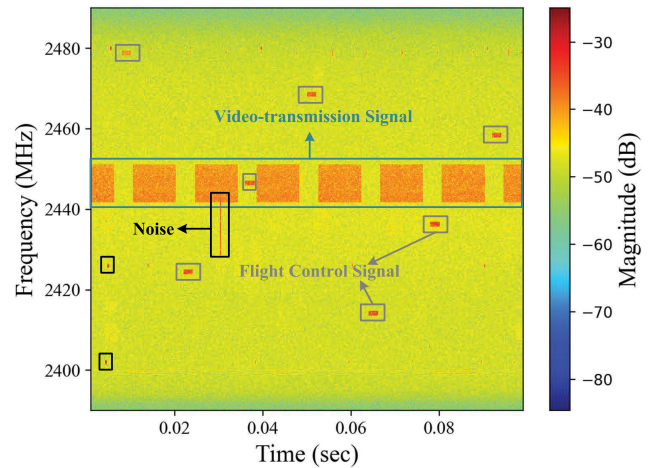


Fig. 1. A time-frequency spectrum of DJI Phantom 4 Pro.

[3], [4], [5], [6], [7], [8]. However, lots of issues caused by abusing drones raise critical concerns on public safety and personal privacy. For example, it was reported that a small drone crashed on the White House lawn in January 2015. In October 2021, the French president's official residence was illegally spied on by two drones. To deal with these threats, geo-fencing and drone-restricted zones are requested to be set up in sensitive areas, such as airports, nuclear facilities, and hospitals, which can protect them from the invasions of registered drones. However, a large number of intruder drones are still unregistered, which brings an urgent requirement to recognize malicious drones.

Passive radio frequency (RF)-based approaches are promising for drone recognition due to their friendliness to urban environments and insensitivity to non-line-of-sight (NLoS) scenarios. As shown in Fig. 1, taking the popular DJI Phantom 4 Pro as a typical example, the communication signal between the drone and the ground controller includes flight control signals (FCS) and video-transmission signals (VTS). To be specific, the FCS refers to the signal carrying flight instructions and flight logs, e.g., speed, altitude, etc. The VTS refers to the signal carrying the video stream from the on-board camera of the drone. Through deploying RF receivers to collect RF signals from the surrounding environment, the communication activities of intruding drones can be effectively surveilled.

Existing RF-based works [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21] primarily focus on the

supervised scenarios, which can identify drones at the assumption of the closed set where the testing signals have to own the same classes as training signals. However, with the explosive increase of drones, it is unrealistic to exhaust all possible classes within a closed set. Therefore, the open set recognition (OSR) for drones has been becoming a pressing matter. The OSR considers a scenario where the unseen/unknown instances during training will appear in testing, which requires the classifier to not only accurately classify the seen/known instances but also effectively cope with those unseen/unknown instances [22], [23].

As the OSR for RF-based drone recognition has not yet been extensively studied, we conclude the challenges that it differs from common OSR methods as follows. (1) **Effective identification for same-brand drones.** Existing OSR methods are generally equipped with convolution-based feature extractors (CFEs). The CFEs perform well in recognition tasks with strong correlations on local features, such as visual images. However, the CFEs cannot handle same-brand drone recognition tasks, as the high similarity of signal modulations leads to indistinguishable local features. Therefore, there is an urgent need to develop drone-oriented feature extractors. (2) **Refined identification for unknown drones.** Different from the OSR commonly applied for image, text, and speech that can be manually reclassified, difficulties in manual distinction necessitate the OSR for RF-based drone recognition not only detecting an unknown signal but also classifying multiple unknown signals into their exact classes. (3) **Insufficient knowledge towards unknown drones.** Some existing OSR methods require the knowledge of unknown classes to determine the optimal bounding thresholds instead of manual thresholds, which can improve the recognition performance. However, considering real-world scenarios, priori knowledge of unknown classes is always unavailable. Therefore, self-adaptive thresholds depending solely on the knowledge of known classes are necessary to be explored.

To address the above-mentioned challenges, we propose a Signal Semantic-based OSR (S3R) method for RF-based drone recognition, which consists of three key modules, i.e., feature generation, semantic construction, and semantic classification. In the feature generation module, short-time Fourier transform (STFT) is employed to generate the time-frequency spectra (TFS), which decouples the target drone signals with the other same-band interference signals. Then, the semantic construction is implemented with a texture extractor (TE) and a position extractor (PE), which respectively learn texture features (bandwidth, time span, etc.) and position features (intervals, frequency points, etc.) from the TFS. The constructed semantics are structurally optimized by reducing the overlap of different-class semantics and enhancing the aggregation of same-class semantics. Finally, a two-stage semantic classification method is formulated. In the first stage, self-adaptive thresholds are obtained by analyzing the outlier known semantics in the closed set, which act as the classification boundaries to detect the unknown semantics. In the second stage, unknown semantics are refinedly recognized via clustering, where the clustering quality analysis is introduced to estimate the most likely number of unknown classes.

Our main contributions are summarized as follows.

- We propose a dilated convolution layers (DCLs)-based TE, which provides multi-level receptive fields to enhance joint awareness of the texture features of FCS and VTS.
- We propose a transformer-based PE that extracts the position features of FCS and VTS from the TFS, which helps in identifying same-brand drones.
- An outlier analysis-based semantic classifier is proposed to determine the self-adaptive bounding thresholds by analyzing the outlier known semantics in the closed set, which act as the classification boundaries to relax the necessity of introducing the open set to our method.
- A semantic augmentation method is proposed to improve the refined recognition of unknown drones, which augments the raw semantics with the knowledge of basic features from the intermediate layers of TE. In this way, the raw semantic space can be expanded to sparsely represent the unknown semantics of the open set.

The rest of this paper is organized as follows. In Section II, the related work is discussed. In Section III, we analyze the signal model in detail and point out some important signal features of drones. In Section IV, the recognition method is proposed. In Section V, we introduce the released dataset briefly. In Section VI, extensive performance evaluation is presented. In Section VII, we conclude this work.

II. RELATED WORK

In this section, we first introduce the basic techniques for drone recognition, especially elucidating the RF-based methods in detail. Next, we introduce the latest research on OSR, especially clarifying the limitations of existing OSR methods.

A. Drone Recognition

The basic techniques for drone recognition include radar, vision, acoustic, and RF [2], [24].

Radar is one of the most classical techniques in aircraft recognition. Existing works mainly used several typical features in radar signals to identify drones, such as radar cross-section [25], micro-doppler signature [26], [27], [28], and radar polarimetry [29]. However, radar-based drone recognition can be less effective due to its Line-of-Sight (LoS) requirement, which cannot be satisfied in crowded urban scenarios. Besides, the high-power instantaneous radiation does harm to the human body, which is also an unavoidable problem.

Vision-based drone recognition is popular due to its intuitive results. Ashraf et al. [30] proposed a two-stage segmentation-based approach employing spatio-temporal attention cues to recognize drones with small sizes. Furthermore, Wang et al. [31] proposed a feature super-resolution-based drone detector with motion information extractor to introduce the scale-invariant features. Although vision-based recognition can be friendly to the urban environment, it also suffers from the LoS requirement and brightness conditions.

Acoustic-based drone recognition is capable of solving the NLoS problem. In [32], Shi et al. achieved drone recognition

and localization by analyzing acoustic signatures generated by drones' motors and rotating propellers. Moreover, in [33], enhanced fiber-optic acoustic sensors for distributed acoustic sensing were used to detect drone sound, which greatly improved the measured sensitivity. However, acoustic-based drone recognition can be unreliable due to the high similarity of sound generated by other electric-powered devices, e.g., electric weed whackers.

RF-based drone recognition mainly refers to recognizing drones by passively monitoring the RF signals emitted by the drone's remote controller or itself. Recently, RF-based drone recognition has been widely used due to its advantages of insensitivity to the NLoS scenarios and friendliness to the urban environment. In [9], Xie et al. used hand-crafted features to recognize drones via a field-programmable gate array, which can identify the number, type, and frequency band of drones in a complex electromagnetic environment. However, the main drawback of this work is that the hand-crafted features of monitoring targets need to be known in advance, including the bandwidth, duration time, and number of occurrences. With the rapid development of machine learning and deep learning, most existing RF-based works employed the learning-based approach to learn RF features automatically for identifying individual emitters from the captured in-phase/quadrature (I/Q) data. Specifically, fast Fourier transform [10], STFT [11], mel-frequency cepstral coefficient [12], wavelet transform [13], empirical mode decomposition [14], etc., are employed to extract the signal features from the I/Q data. Then, classifiers are designed to classify these signal features based on the support vector machine [15], fully-connected network [10], convolution neural network [16], long short-term memory network [17], and so on. Considering the requirement of low complexity in actual system deployment, Cai et al. [18] proposed a lightweight model including multi-scale convolution blocks, which achieves promising results at low signal-to-noise ratios in terms of both accuracy and efficiency. Besides, more and more consumer-grade drones are equipped with Wi-Fi interfaces to communicate with users' smartphones. To address this issue, Alipour-Fanid et al. [19] proposed a machine learning-based framework for drone recognition over encrypted Wi-Fi traffic, which exploits the RF features derived from packet size and inter-arrival time of traffic data. In the defense of autopilot drones, Deng et al. [20] developed a proactive detection system named Dr. Defender, which uses the representative features of the unique motions of drones from the channel state information of WiFi. Moreover, Lin et al. [21] considered autopilot drone detection in an outdoor scenario, where the motions of drones represented in the synchronization signal blocks of the fifth-generation (5G) signals are studied. However, all the above-mentioned RF-based recognition methods belong to the supervised methods, which follow the assumption of the closed set and thus cannot identify unknown classes in the open set.

B. OSR

To our best knowledge, there has been few research on OSR for RF-based drone recognition. Here, we summarize commonly-used methods of OSR as follows.

In most neural network (NN)-based models, SoftMax function is used after the last fully-connected layer. However, it assigns both known and unknown classes to a uniform feature space, leading to the closed set nature inherently. To handle this issue, Soltani et al. [34] introduced the strategy of thresholding SoftMax probability. The task of detecting an unknown drone is accomplished by a pair of statistically derived thresholds, i.e., the SoftMax probability and the correctly classified ratio of a multi-NN model. In [35], OpenMax layer is connected after the penultimate layer of the NN model to estimate the probability of an input from being unknown, where the separate Weibull distribution is fitted for each known class based on the training samples. In [36], an uncertainty-inspired open set (UIOS) model was proposed to obtain the feature evidence by using the Softplus activation function. The feature evidence is then parameterized to a Dirichlet distribution. Based on the Dirichlet distribution, the uncertainty score of each input can be calculated to determine whether the input is known or unknown. Besides, Li et al. [37] proposed a multi-modal marginal prototype framework for modulated RF signal open-set learning, where the spatial distribution of each class is optimized by adjusting reciprocal point learning and prototype learning jointly. Then, a generative adversarial network (GAN)-based unknown sample generation strategy was proposed to enable a better adaptation to the unknown world. However, the above-mentioned methods can only reject unknown samples but fail to identify multiple unknown instances as their true classes. Essentially, they fall into the scope of anomaly detection.

To achieve the refined recognition for unknown classes, a signal recognition and reconstruction convolutional neural network (SR2CNN) [38] was proposed to recognize the modulation signals. Specifically, an auto-encoder is applied to map representations of input signals in a feature space. Then, the classification is implemented by measuring the Mahalanobis distance from the input sample to the center of each class in the feature space, which are compared with preset bounding thresholds. In [39], Han et al. employed the Euclidean distance as the distance measurement instead of the Mahalanobis distance to improve the recognition performance of unknown jamming patterns. Nevertheless, although the above-mentioned methods can achieve refined unknown recognition, they have to introduce unknown data from the open set to determine the bounding thresholds. This means that these methods cannot work in scenarios where unknown data are unavailable.

III. SIGNAL MODEL AND ANALYSIS

The raw RF signals of drones are typically coupled in the time domain, which can be expressed as

$$x(t) = x_v(t) + x_c(t) + x_i(t) + n(t), \quad (1)$$

where $t = 1, 2, \dots, T$, and T is the sampling length. Here, $x_v(t)$ is the downlink VTS sent from the drone to the controller, modulated by orthogonal frequency division multiplexing (OFDM), whose general mathematical model can

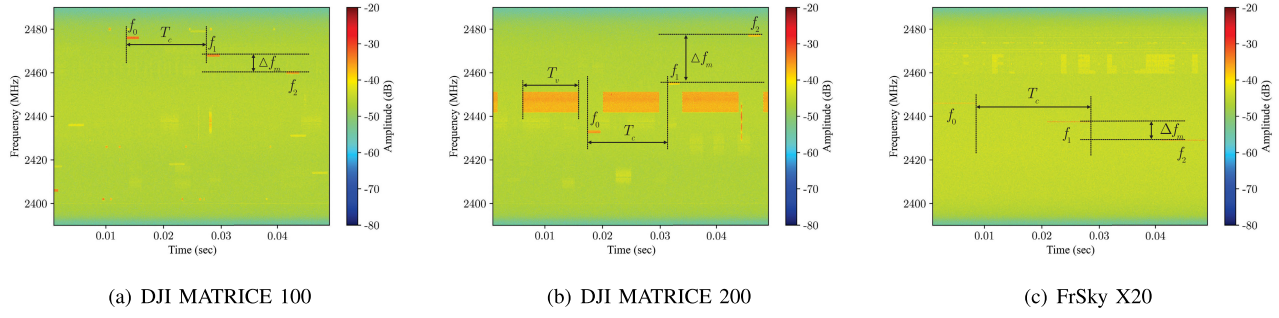


Fig. 2. An illustration of the difference in RF characteristics among three typical drones.

be denoted as [9]

$$x_v(t) = \begin{cases} \sum_{n=1}^{N_v-1} b_n \text{rect}(t - \Delta t - \frac{T_v}{2}) e^{j2\pi \frac{n}{T_v}(t - \Delta t)}, & t \in [\Delta t, \Delta t + T_v], \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where N_v represents the number of OFDM subcarriers, T_v denotes the duration time of the OFDM symbol, b_n ($n = 0, 1, \dots, N_v - 1$) denotes the symbol allocated to each subcarrier, and Δt is the sampling interval. $\text{rect}(\cdot)$ is a rectangular function, where $\text{rect}(x) = 1$ when $|x| \leq T_v/2$. $x_c(t)$ is the uplink FCS between the drone and the controller, modulated by frequency-hopping spread spectrum (FHSS), which can be expressed as

$$x_c(t) = s(t) \cdot A \sum_{m=0}^{M_c-1} W_T(t - mT_c) \cos(2\pi f_m(t - mT_c) + \phi_m), \quad (3)$$

where $s(t)$ is the modulated control signal, M_c denotes the number of frequency-hopping points, A represents the modulation amplitude, W_T is the rectangular window, and T_c is the frequency-hopping period. $\mathcal{F} = \{f_0, \dots, f_m, \dots, f_{M_c-1}\}$ is a set of frequency points. For most drones, the frequency difference $\Delta f_m = |f_{m_i} - f_{m_j}|$ are regular values, where $m_i, m_j \in \{0, 1, \dots, M_c - 1\}$. $\{\phi_m\}_{m=0,1,\dots,M_c-1}$ is a set of initial phases. $x_i(t)$ denotes the interference signal, mainly including Bluetooth and Wi-Fi of other devices. $n(t)$ is the inherent noise caused by the RF receiver, e.g., Gaussian white noise.

To begin up recognizing drones, some key RF characteristics of common drones are summarized as follows.

- The FCS generally have a bandwidth of smaller than 5MHz and a duration of less than 5ms, whereas the VTS typically have a bandwidth of 10MHz or 20MHz and a duration of greater than 10ms.
- The power of the VTS is higher than the FCS. However, while the video transmission can be voluntarily cut off by the user in some scenarios, the recognition focusing on the FCS is non-negligible but more challenging.

- Various modulation parameters result in different RF characteristics of drones, i.e., the modulation amplitude A , the rectangular window W_T , the duration time of the OFDM symbol T_v , the frequency-hopping period T_c , the frequency difference Δf_m , and the frequency points \mathcal{F} . Considering the local correlations on the TFS, A , T_v , and W_T can be regarded as **texture features**. Besides, T_c , Δf_m , and \mathcal{F} can be considered as **position features** due to their global correlations on the TFS.

For the recognition of different-brand drones, the above-mentioned modulation parameters are always important characteristics. However, for the recognition of same-brand drones, sometimes only the position features differ, which increases the recognition difficulty due to the lack of the texture features.

Fig. 2 illustrates the difference in RF characteristics between the same-brand drones and the different-brand drones with a typical example. Specifically, DJI MATRICE 100 and DJI MATRICE 200 belong to the same brand of DJI MATRICE, while FrSky X20 belongs to another brand. Notice that the VTS of DJI MATRICE 100 are cut off, so the difference between DJI MATRICE 100 and DJI MATRICE 200 is only in the position features, i.e., T_c , Δf_m , and \mathcal{F} marked in Fig. 2.

Based on the above analysis of drone signals, it is worth noting that the key features of the TFS are fundamentally different from common visual images, which makes common OSR methods for visual images ineffective in dealing with the TFS.

IV. METHODOLOGIES

The overview of the proposed S3R method is illustrated in Fig. 3, consisting of three key modules, i.e., feature generation, semantic construction, and semantic classification, which are detailedly introduced as follows.

A. Feature Generation

To perform real-time and noise-insensitive drone recognition, STFT is employed to decouple target drone signals with interference signals in the time-frequency domain [40]. The STFT first divides the captured signal into several sub-segments with the same length and then performs fast Fourier transform on each sub-segment independently. In this way, the signal information in both time domain and frequency domain can be jointly represented in a 2-dimensional

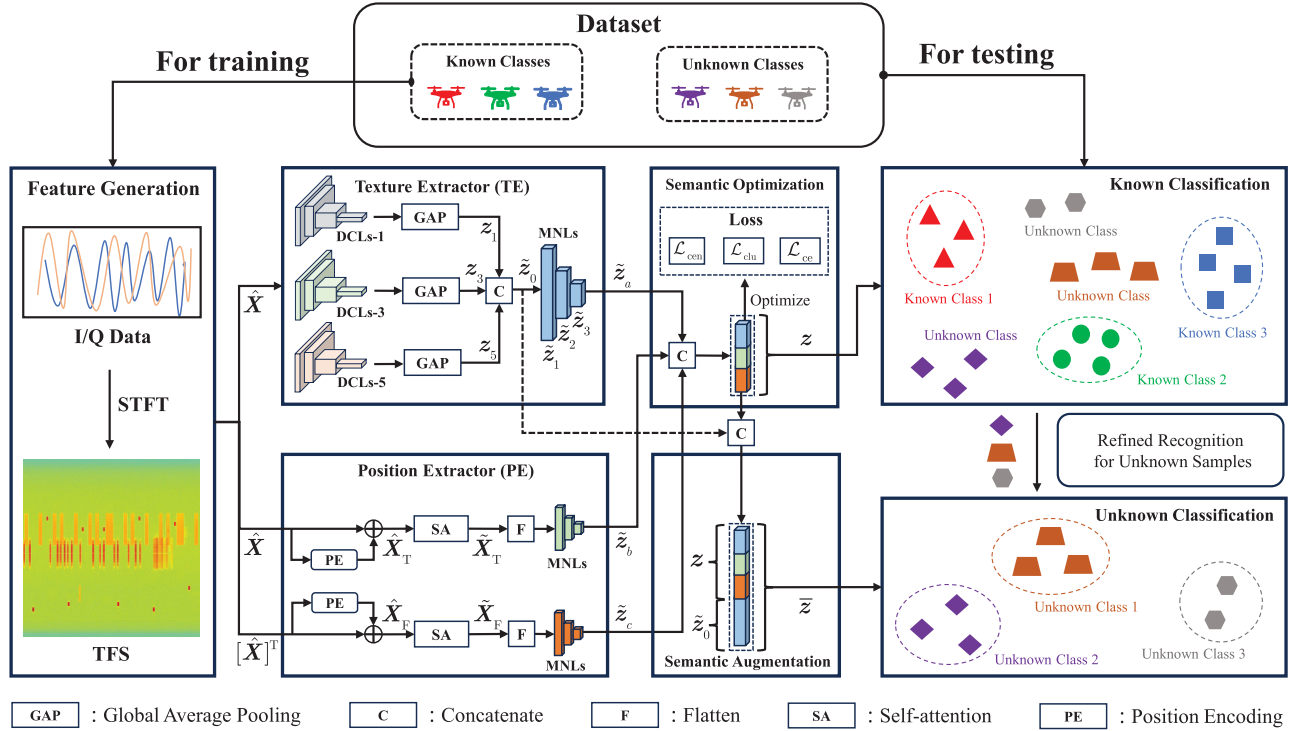


Fig. 3. Overview of the proposed S3R method for drone recognition.

(2-D) matrix. Finally, by taking the amplitude and logarithmic scaling, the power spectrum of the captured signal can be explicitly represented as

$$X(n, w) = 20 \log_{10} \left| \sum_{t=1}^W x(t + nR_W) g(t) \cdot e^{(-j2\pi \frac{w}{W} t)} \right|, \quad (4)$$

where R_W is the sliding length of the short-time analysis window, and $g(t)$ is the Hamming window function to mitigate the spectrum leakage. $w \in \mathcal{W} = \{1, 2, \dots, W\}$ with W being the length of the short-time analysis window. $n \in \mathcal{N} = \{1, 2, \dots, N\}$ with N being the number of the short-time analysis window. $|\cdot|$ indicates the modulus of the complex number.

The high mobility of drones results in large fluctuations in channel state, which makes the signal power of drones close to the RF receiver much greater than that of remote drones. To maintain a fair comparison, the signal power spectrum is normalized as

$$\hat{X}(n, w) = \frac{X(n, w) - \min[X]}{\max[X] - \min[X]}, \quad (5)$$

where $\max[\cdot]$ and $\min[\cdot]$ represent the maximum and minimum of the elements in the inside matrix, respectively.

B. Semantic Construction

In this subsection, we first propose a TE and a PE to construct the signal semantics for representing drone TFS, which are then optimized to make the semantics distinguishable.

1) *Texture Extractor (TE)*: According to the analysis in Section III, TE is designed to learn the texture features, focusing on bandwidths and time duration of different drones. Due to the size differences in spectral representations between

FCS and VTS, TE is embedded with three DCLs whose dilated setting is c ($c = 1, 3, 5$), providing a multi-level receptive field to facilitate the texture feature learning. The i -th feature map in DCLs- c can be expressed as

$$\mathbf{Z}_i^c = \text{MaxPool}(\text{ReLU}(\mathbf{W}_i^c * \mathbf{Z}_{i-1}^c + \mathbf{b}_i^c)), i = 1, 2, \dots, I_{\max}, \quad (6)$$

where $\mathbf{Z}_0^c = \hat{\mathbf{X}}$, \mathbf{W}_i^c and \mathbf{b}_i^c denote the learnable parameters, $\text{MaxPool}(\cdot)$ is the operation of MaxPooling, $*$ denotes the convolution operation, $\text{ReLU}(\cdot)$ is the ReLU activation function, and I_{\max} is the number of feature maps of DCLs- c . The 2-D outputs of each DCLs- c are respectively down-sampled into one-dimensional (1-D) vectors by using the operation of the global average pooling, denoted as

$$\mathbf{z}_c = \text{GlobalAvgPool}(\mathbf{Z}_{I_{\max}}^c), c = 1, 3, 5, \quad (7)$$

where $\text{GlobalAvgPool}(\cdot) : \mathbb{R}^{d_1 \times d_2} \rightarrow \mathbb{R}^1$ represents the operation of outputting the mean of all elements of the input 2-D feature map with the size of $d_1 \times d_2$. The 1-D outputs of three DCLs are further concatenated to be the input of the multiple non-linear layers (MNLs) as

$$\tilde{\mathbf{z}}_0 = \mathbf{z}_1 \oplus \mathbf{z}_3 \oplus \mathbf{z}_5, \quad (8)$$

where \oplus denotes the operation of concatenating two 1-D inputs, that is, $\mathbb{R}^{d_1} \oplus \mathbb{R}^{d_2} \rightarrow \mathbb{R}^{d_1+d_2}$ with d_1 and d_2 being the size of these two inputs, respectively. Given the MNLs with J_{\max} layers, the MNLs can be expressed as

$$\text{MNLs}(\tilde{\mathbf{z}}_0) = \phi_{J_{\max}-1} \left(\phi_{J_{\max}-2} \left(\dots \phi_1 (\phi_0(\tilde{\mathbf{z}}_0)) \right) \right), \quad (9)$$

where

$$\phi_j(\tilde{\mathbf{z}}_j) = \text{ReLU}(\mathbf{w}_j * \tilde{\mathbf{z}}_j), j = 0, 1, \dots, J_{\max} - 1, \quad (10)$$

where w_j are the learnable parameters of the j -th layer of the MNLs. Therefore, the texture semantics can be recursively constructed by the MNLs, denoted as $\tilde{z}_a = \text{MNLs}(\tilde{z}_0)$.

2) *Position Extractor (PE)*: Due to the translation invariance as explained in Appendix A, the DCLs-based TE is not capable of identifying the same-brand drones. In other words, the position features including the absolute position features \mathcal{F} and the relative position features $\Delta f_m, T_c$ cannot be efficiently extracted by TE. Therefore, PE is designed to effectively extract the position features. Different from TE using the convolution structures, PE utilizes the positional encoding added in the inputs to emphatically model the position features, which is explained in Appendix B. Given the 2-D input \hat{X} , the position encoding of $\hat{X}(n, w)$ can be written as

$$P(n, w; \hat{X}) = \begin{cases} \sin(\omega_w n), & \text{if } w \% 2 = 1, \\ \cos(\omega_w n), & \text{if } w \% 2 = 0, \end{cases} \quad (11)$$

where $\%$ denotes the operation of calculating the remainder. ω_w is the hand-crafted frequency, and typically, $\omega_w = 10000^{-\frac{2w}{W}}$. To highlight the position features, the corresponding position encoding is added to the input as well as its transposed matrix, respectively, which is expressed as

$$\hat{X}_T(n, w) = \hat{X}(n, w) + P(n, w; \hat{X}), \quad (12)$$

$$\hat{X}_F(w, n) = [\hat{X}]^T(w, n) + P(w, n; [\hat{X}]^T), \quad (13)$$

where $[\cdot]^T$ denotes the operation of matrix transpose. In this way, \hat{X}_T can reflect the position features in the time domain, and \hat{X}_F can reflect the position features in the frequency domain.

Then, a set of self-attention (SA) modules is employed to extract the latent position features, which performs well in learning the long-range dependencies and interactions in sequential data [41]. The SA module is particular of three kinds of learnable matrices, namely, query matrix Q , key matrix K , and value matrix V . Specifically, Q queries the features related to drones among all input features, K emphasizes the correlation of the drone features with each other, and V assigns weights to all input features by calculating the similarity between the features represented by Q and K . The extraction process of the position features for any input Y_0 can be expressed as

$$\text{Atten}(Y_0) = \varphi_{L_{\max}-1} \left(\varphi_{L_{\max}-2} \left(\dots \left(\varphi_1(\varphi_0(Y_0)) \right) \right) \right), \quad (14)$$

where L_{\max} is the number of the SA modules, and the l -th SA module φ_l can be denoted as

$$\varphi_l(Y_l) = \text{softmax} \left(\frac{Y_l Q_l [K_l]^T [Y_l]^T}{\sqrt{\gamma_l}} \right) Y_l V_l, \quad (15)$$

where Y_l is the input of the l -th SA module, $l = 0, 1, \dots, L_{\max}-1$, and $\text{softmax}(\cdot)$ denotes the SoftMax activation function. $Q_l, K_l, V_l \in \mathbb{R}^{n_l \times \gamma_l}$ are learnable parameters of the l -th SA module, where n_l and γ_l are the input dimension and the mapping dimension of the l -th SA module, respectively. The outputs of the set of SA modules can be expressed as

$$\tilde{X}_T = \text{Atten}(\hat{X}_T), \quad \tilde{X}_F = \text{Atten}(\hat{X}_F). \quad (16)$$

Then, \tilde{X}_T and \tilde{X}_F are respectively flattened as 1-D vectors, which are further mapped through the MNLs, denoted as

$$\tilde{z}_b = \text{MNLs}(\text{Flatten}(\tilde{X}_T)), \quad (17)$$

$$\tilde{z}_c = \text{MNLs}(\text{Flatten}(\tilde{X}_F)) \quad (18)$$

where $\text{Flatten}(\cdot) : \mathbb{R}^{d_1 \times d_2} \rightarrow \mathbb{R}^{d_1 d_2}$ is the operation of reshaping the 2-D matrix into the 1-D matrix.

3) *Semantic Optimization*: Since TE and PE respectively extract the texture and position features from the TFS, the signal semantics are finally obtained by fusing these two features to improve the distinguishable representations, denoted as

$$z = \text{MNLs}(\tilde{z}_a \oplus \tilde{z}_b \oplus \tilde{z}_c), \quad (19)$$

where $z \in \mathbb{R}^D$, and D is the dimension of the semantic space.

To make z distinguishable for different drones, it is necessary to optimize the learnable parameters. Let θ represent all parameters of TE and PE, including W, b, w, Q, K and V , and let $E(\cdot)$ represent the TFS normalization and the semantic construction process. First, the center loss is introduced to make the same-class semantics retain close by. Given a training batch $\mathcal{M} = \{(X_1, y_1), \dots, (X_i, y_i), \dots, (X_M, y_M)\}$ with batch size M , where $y_i \in \mathcal{K}$ with $\mathcal{K} = \{1, 2, \dots, K\}$ being the class set of the closed set, the center loss is denoted as

$$\mathcal{L}_{\text{cen}}(\theta; \mathcal{M}) = \sum_{X_i \in \mathcal{M}} \|E(X_i; \theta) - \mu_{y_i}\|_2, \quad (20)$$

where μ_{y_i} is the semantic center of the known class y_i , and $\|\cdot\|_2$ represents the operation of calculating the 2-norm. Given the learning rate α , μ_{y_i} can be updated by $\mu_{y_i} \leftarrow \mu_{y_i} - \alpha \Delta \mu_{y_i}$ in each batch, and $\Delta \mu_{y_i}$ can be obtained via

$$\Delta \mu_{y_i} = \begin{cases} 0, & \text{if } \sum_{j=1}^M \delta(y_j = y_i) = 0, \\ \frac{\sum_{j=1}^M \delta(y_j = y_i) (E(X_j; \theta) - \mu_{y_i})}{\sum_{j=1}^M \delta(y_j = y_i)}, & \text{otherwise,} \end{cases} \quad (21)$$

where $\delta(x) = 1$ if the inside condition x is true, and $\delta(x) = 0$ otherwise.

Second, the different-class semantics should be placed far apart in the semantic space, which can be achieved via the clustering loss

$$\mathcal{L}_{\text{clu}}(\theta; \mathcal{M}) = \sum_{X_i \in \mathcal{M}} \sum_{k \in \mathcal{K}} \max\{0, \vartheta - \|E(X_i; \theta) - \mu_k\|_2\}, \quad (22)$$

where ϑ is the margin to divide the different-class semantics.

Third, the constructed semantics can be directly used for the supervised classification, and thus we employ the cross entropy loss, denoted as

$$\mathcal{L}_{\text{ce}}(\theta; \mathcal{M}) = -\frac{1}{M} \sum_{X_i \in \mathcal{M}} y_i \log(P(X_i; \theta)), \quad (23)$$

where $P(X_i; \theta)$ is the predicted probability of X_i belonging to the class y_i . Specifically, the predicted probability can be obtained by deploying an MNLs-based classifier with the semantic as the input and the output activated by the SoftMax function.

Finally, θ can be optimized by minimizing the sum of the above-mentioned three loss functions via stochastic gradient descent (SGD), denoted as

$$\theta \leftarrow \theta - \alpha \frac{\partial(\eta_1 \mathcal{L}_{\text{cen}}(\theta) + \eta_2 \mathcal{L}_{\text{clu}}(\theta) + \eta_3 \mathcal{L}_{\text{ce}}(\theta))}{\partial \theta}, \quad (24)$$

where η_1, η_2, η_3 are weight parameters.

C. Semantic Classification

After training the proposed feature extractors, the semantics of the closed set with class label k can be gathered in the semantic space \mathbb{R}^D , resulting in a class domain denoted as

$$R_k := \{z \in \mathbb{R}^D \mid \text{dist}(z, \mu_k) < \Theta_k\}, \quad (25)$$

where Θ_k is a core bounding threshold, and

$$\text{dist}(z, \mu_k) = \sqrt{(z - \mu_k)^T \Sigma_k^{-1} (z - \mu_k)} \quad (26)$$

is the Mahalanobis distance between z and μ_k . Here, Σ_k^{-1} is the inverse of the covariance matrix of the semantics belonging to the class k .

Intuitively, a testing semantic z_t should be recognized as a known class if $z_t \in \{R_k\}_{k \in \mathcal{K}}$ holds. Otherwise, z_t is considered as an unknown class, since z_t is not included in the domain of any known class. Therefore, the semantic classification aims to search a set of $\{\Theta_k\}_{k \in \mathcal{K}}$, which can optimally divide the semantic space so that each class domain contains only one class of semantics as much as possible. In [38] and [39], an optimal setting of Θ_k^* is found by using the grid search, which aims to balance the recognition performance of the classifier on both known and unknown testing semantics, denoted as

$$\Theta_k^* = \arg \max_{\Theta_k \in (0, +\infty)} \beta \Phi(\mathcal{X}_K) + (1 - \beta) \Phi(\mathcal{X}_U), \quad (27)$$

where \mathcal{X}_K and \mathcal{X}_U are the testing set of known samples and unknown samples, respectively. $\beta \in (0, 1)$ is a weight parameter. $\Phi(\cdot)$ is the evaluation metric of the semantic classifier, which can be accuracy, precision, or a self-defined indicator.

Considering the fact that \mathcal{X}_U is always unavailable, it is necessary to develop a semantic classification method that only relies on \mathcal{X}_K to determine Θ_k^* . Inspired by the important finding [38] that the structural distribution of known semantics most likely follows the Gaussian distribution when \mathcal{L}_{cen} is used to gather the same-class semantics, an outlier analysis-based classifier is proposed to obtain the setting of Θ_k^* with the closed set. Specifically, the set of distances from training semantics with class label k to μ_k can be counted as

$$\mathcal{P}_k = \{\text{dist}(z_1, \mu_k), \text{dist}(z_2, \mu_k), \dots, \text{dist}(z_{N_k}, \mu_k)\}, \quad (28)$$

where N_k is the number of training samples belonging to class k . Then, \mathcal{P}_k is resorted in descending order as $\hat{\mathcal{P}}_k =$

$\{\varrho_k^1, \dots, \varrho_k^{\hat{n}}, \dots, \varrho_k^{N_k}\}$ where $\varrho_k^{\hat{n}}$ is the \hat{n} -th element in $\hat{\mathcal{P}}_k$ and $\varrho_k^{\hat{n}-1} \geq \varrho_k^{\hat{n}}$, $\hat{n} = 2, \dots, N_k$. Finally, following the 3σ rule in the statistical theory, $\hat{\mathcal{P}}_k$ is traversed to find the first element that is not an outlier as Θ_k^* , which can be expressed as

$$\Theta_k^* = \arg \min_{\varrho_k^{\hat{n}}} \hat{n} \cdot \delta(\varrho_k^{\hat{n}} \geq 3\sigma_k), \quad \hat{n} = 1, 2, \dots, N_k, \quad (29)$$

where σ_k is the standard deviation of \mathcal{P}_k .

1) *Known Classification*: When $z_t \in \{R_k\}_{k \in \mathcal{K}}$ holds, z_t is recognized as one of the known classes, whose exact class \hat{y}_t is denoted as

$$\hat{y}_t = \arg \min_{k \in \mathcal{K}} \text{dist}(z_t, \mu_k). \quad (30)$$

2) *Unknown Classification*: When $z_t \notin \{R_k\}_{k \in \mathcal{K}}$ holds, z_t is recognized as an unknown class, which will be further refinedly identified. Since the raw semantics z are constructed in a supervised manner to recognize the closed set, it is of great importance to augment z to address the open set. Inspired by the work [42], raw semantics are augmented by fusing abstractive features from the intermediate layers of TE, denoted as $\bar{z} = z \oplus \bar{z}_0$. The theoretical basis of this augmentation method is that the MNLs are trained to ignore some abstractive features (i.e., vague and fragmented features represented by convolution layers) irrelevant to the closed set, while these abstractive features are beneficial to the representations of unknown classes. In the proposed S3R method, these abstractive features are fully preserved at the outputs of DCLs.

Before refinedly classifying the unknown semantics, let $\mathcal{D} = \{\bar{z}_1, \bar{z}_2, \dots, \bar{z}_{N_U}\}$ denote the set of unknown semantics, where N_U denotes the number of unknown semantics to be classified. Assume the potential number of unknown classes is u , the problem of classifying unknown semantics is transformed as dividing \mathcal{D} into several groups, i.e., $\mathcal{C}_u = \{C_1, C_2, \dots, C_u\}$.

Case-A: $u = 1$. If all the unknown semantics are clustered together, all testing unknown samples are recognized as the same unknown class. By calculating the outlier-based threshold Θ_U^* for \mathcal{D} according to Eq. (29), the decision of $u = 1$ is made when $\Theta_U^* < \max\{\Theta_k^*\}_{k \in \mathcal{K}}$ holds.

Case-B: $u \neq 1$. In this case, we consider $u = 2, 3, \dots, N_U$. Given an exact number of u , the semantic division can be optimized with the following objective as

$$\underbrace{C_1, \dots, C_i, \dots, C_u}_{\mathcal{C}_u} = \arg \min \sum_{i=1}^u \sum_{\bar{z} \in C_i} \|\bar{z} - \bar{\mu}_i\|_2^2, \quad (31)$$

where $\bar{\mu}_i$ is the center of C_i . Eq. (31) demonstrates that \mathcal{C}_u can be optimally found with the objective of minimizing semantic differences within the same group, that is, expecting similar semantics to be classified into the same group. The optimization of Eq. (31) can be achieved via clustering algorithms, e.g., k-means. Introducing an evaluation function $\Psi(\cdot)$, the quality of the semantic division can be expressed as $\Psi(\mathcal{C}_u)$. $\Psi(\mathcal{C}_u)$ respect to u is likely to be an extreme point when u reaches the true number of unknown classes. This is because the evaluation function comprehensively examines the local aggregation and global sparsity of the semantic space, which

is consistent with our goal of the semantic optimization, that is, clustering same-class semantics and separating different-class semantics. Therefore, the number of unknown classes can be inferred as

$$u^* = \arg \max_{u=2,3,\dots,N_U} |\Psi(C_{u-1}) + \Psi(C_{u+1}) - 2\Psi(C_u)|, \quad (32)$$

where $|\cdot|$ denotes taking the absolute value. The clustering evaluation $\Psi(\cdot)$ is implemented via the Silhouette Coefficient, denoted as

$$\Psi(C_u) = \sum_{t=1}^{N_U} \frac{a(\bar{z}_t) - b(\bar{z}_t)}{\max\{a(\bar{z}_t), b(\bar{z}_t)\}}, \quad (33)$$

where $\bar{z}_t \in \mathcal{D}$. $a(\bar{z}_t)$ and $b(\bar{z}_t)$ are respectively expressed as

$$a(\bar{z}_t) = \frac{\sum_{1 \leq t' \leq |C_i|} \text{dist}(\bar{z}_t, \bar{z}_{t'})}{|C_i| - 1}, \quad \bar{z}_t, \bar{z}_{t'} \in C_i, \quad (34)$$

$$b(\bar{z}_t) = \min_{C_j} \text{dist}(\bar{z}_t, C_j), \quad \bar{z}_t \in C_i, C_j \neq C_i, \quad (35)$$

where $|C_i|$ denotes the number of samples in the i -th semantic cluster. With u^* determined in Eq. (32), the recognition of \mathcal{D} can be finally obtained as C_{u^*} by solving the optimization problem in Eq. (31) again.

We summarize the proposed S3R method in **Algorithm 1**.

V. DATASET

We provide a real-world spectrogram dataset, denominated as *DroneRFb-Spectra*, which records 24 classes of RF signals captured by the Universal Software Radio Peripheral (USRP) device. Specifically, the dataset captures 10 classes outdoors, where drones are operated 20~150 meters away from the receiver,¹ as shown in Fig. 4. Besides, the other 14 classes are captured indoors, where drones are in hovering mode with 5 meters away from the receiver. The 24 classes cover 7 brands, namely, DJI, VBar, FrSky, Futaba, Taranis, RadioLink, and Skydroid. The RF monitoring covers three Industrial Scientific Medical (ISM) bands, i.e., 915MHz, 2.4GHz, and 5.8GHz with a bandwidth of 100MHz. We have sliced the collected I/Q data into small segments with a single length of 50ms, which are further transformed into the TFS with a downsampled size of 512×512 . The detailed introductions of each class are shown in Table I. For more introduction of the dataset, please check the documentation under the released access addresses.²

VI. EXPERIMENT

A. Experimental Setting

The setup of building and training the proposed model is shown in Table II. Besides, we employ the k-means clustering algorithm [43] to solve the problem in Eq. (31).

The recognition model is trained on the NVIDIA GeForce RTX 3090 with a maximum epoch of 250. The curve of loss

¹We set the receiving gain of the RF receiver as 50dB, equipped with omnidirectional antennas providing an additional gain of 3dB. Besides, we did not use any power amplifier, which limited the effective detection range of the RF receiver to only 200 meters.

²IEEE DataPort, doi: 10.21227/wv7h-sv64.

Algorithm 1 Proposed S3R for RF-Based Drone Recognition

```

1: Input: Training data  $\{(X_i, y_i)\}_{i=1,2,\dots}$  with certain known
   classes  $\mathcal{K} = \{1, 2, \dots, K\}$ , extractor  $E$ , and training
   hype-parameters  $\alpha, \vartheta, \eta_1, \eta_2, \eta_3$ .
2: Randomly Initialize  $\{\mu_k\}_{k \in \mathcal{K}}$  and  $\theta$ .
3: Stage-I: Semantic Construction
4: while  $\theta$  do not converge do
5:   Randomly select a mini-batch from training data.
6:   Calculate the semantics as  $z = E(X; \theta)$ .
7:   Calculate  $\mathcal{L}_{\text{cen}}, \mathcal{L}_{\text{clu}}, \mathcal{L}_{\text{ce}}$  according to (20), (22) (23),
      respectively.
8:   Update  $\theta$  according to (24).
9:   Update  $\{\mu_k\}_{k \in \mathcal{K}}$  by  $\mu_k \leftarrow \mu_k - \alpha \Delta \mu_k$ .
10: end while
11: Stage-II: Semantic Classification
12: Calculate  $\{\Theta_k^*\}_{k \in \mathcal{K}}$  according to (29).
13: Obtain  $\{R_k\}_{k \in \mathcal{K}}$  according to (25).
14: Set an empty buffer  $\mathcal{D}$ .
15: for  $t = 1, 2, \dots$  do
16:   if  $z_t \in \{\Theta_k^*\}_{k \in \mathcal{K}}$  then
17:      $z_t$  is classified as a known class with its predicted
       label  $\hat{y}_t = \arg \min_{k \in \mathcal{K}} \text{dist}(z_t, \mu_k)$ .
18:   else
19:      $z_t$  belongs to unknown classes, which is put into  $\mathcal{D}$ .
20:     while  $|\mathcal{D}|$  reaches  $N_U$  do
21:       Calculate  $\Theta_U^*$  of  $\mathcal{D}$ .
22:       if  $\Theta_U^* \leq \max\{\Theta_k^*\}_{k \in \mathcal{K}}$  then
23:         All samples in  $\mathcal{D}$  are recognized as the same
           unknown classes, i.e.,  $u = 1$ .
24:       else
25:         for  $u = 2, 3, \dots, N_U$  do
26:           Divide  $\mathcal{D}$  with  $u$  into  $C_u$  according to (31).
27:           Evaluate quality of  $C_u$  as  $\Psi(C_u)$ .
28:         end for
29:         Obtain the number of unknown classes as  $u^*$ 
           according to (32).
30:       end if
31:       Implement clustering algorithms again with  $u^*$  for
          $\mathcal{D}$  and output the unknown classification groups
          $\{C_1, C_2, \dots, C_{u^*}\}$ .
32:     end while
33:   end if
34: end for
35: Output:  $\hat{y}_t \in \mathcal{K}$  or  $\hat{y}_t \in \{C_1, C_2, \dots, C_{u^*}\}$ .

```

with respect to the training epoch is shown in Fig. 5, where the losses of clustering and cross entropy decrease quickly while the center loss decreases slowly. Overall, by adjusting the values of η_1, η_2 , and η_3 as shown in Table II, the proposed model can converge efficiently.

B. Evaluation Indicator

Instead of the commonly-used *accuracy*, the other four indicators are introduced to comprehensively evaluate the proposed method, namely, *true known rate* (TKR), *true unknown rate* (TUR), *known precision* (KP), and *unknown precision*



(a) RF receiver



(b) Monitoring scenario outdoors



(c) Flying drone at 150 meters away

Fig. 4. Experimental and signal collection scenarios.

TABLE I
THE CLASSES OF THE RELEASED DATASET

Index	Class	Composition		Number of Samples
		FCS	VTS	
A	Background [†]	No	No	689
B	DJI Phantom 3	Yes	No	689
C	DJI Phantom 4 Pro	Yes	Yes	710
D	DJI MATRICE 200	Yes	Yes	767
E	DJI MATRICE 100	Yes	No	734
F	DJI Air 2S	Yes	Yes	761
G	DJI Mini 3 Pro	Yes	Yes	719
H	DJI Inspire 2	Yes	Yes	734
I	DJI Mavic Pro	Yes	Yes	347
J	DJI Mini 2	Yes	Yes	743
K	DJI Mavic 3	Yes	Yes	479
L	DJI MATRICE 300	Yes	Yes	296
M	DJI Phantom 4 Pro RTK	Yes	Yes	578
N	DJI MATRICE 30T	Yes	Yes	392
O	DJI AVATA	Yes	Yes	572
P	DJI DIY [‡]	Yes	Yes	572
Q	DJI MATRICE 600 Pro	Yes	Yes	554
R	VBar	Yes	No	584
S	FrSky X20	Yes	No	590
T	Futaba T16IZ	Yes	No	569
U	Taranis Plus	Yes	Yes	614
V	RadioLink AT9S	Yes	No	575
W	Futaba T14SG	Yes	No	584
X	Skydroid	Yes	No	608

[†] The background class refers to the captured Wi-Fi and Bluetooth signals in the urban environment. The Wi-Fi device is H3C router (EWP-WA5320), and the Bluetooth device is a Sony audio (SRS-XB20).

[‡] The DJI DIY class refers to an amateur device assembled with the DJI communication modules in a do-it-yourself (DIY) manner.

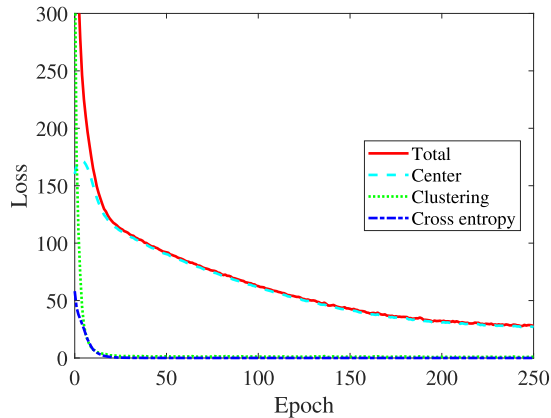


Fig. 5. Training loss with respect to the epoch.

(UP), which are defined as

$$\text{TKR} = \frac{TK}{TK + FU}, \quad \text{TUR} = \frac{TU}{TU + FK},$$

TABLE II
TRAINING HYPE-PARAMETERS

Hype-parameter	Default Setup
Sampling Rate	100 MS/s
T	5×10^6 (50ms)
W	2048
R_W	1024
I_{\max}	6
J_{\max}	3
L_{\max}	3
γ_l	256
n_l	64
α	1×10^{-4}
M	48
ϑ	8
η_1	5×10^{-3}
η_2	1
η_3	0.1
D	128
Optimizer	SGD
Max Epoch	250

$$\text{KP} = \frac{CK}{TK + FK}, \quad \text{UP} = \frac{CU}{TU + FU}, \quad (36)$$

where TK , TU , FK , FU respectively denote the number of testing samples recognized as *truly known*, *truly unknown*, *falsely known*, *falsely unknown*. *Correctly known* (CK) represents the total number of testing known samples that are correctly classified to their exact known classes. *Correctly unknown* (CU) represents the total number of same-class unknown samples classified together, only when the same-class unknown samples can dominate their groups (at least 50%).

C. Comparison Methods

To show the superiority of the proposed method, the compared methods with their goals and data requirements are concluded in Table III. Specifically, SR2CNN [38], OpenMax [35], and UIOS [36] belong to the state-of-art OSR methods, whereas Isolation Forest (IF) [44], One-class SVM [45], and Local Outlier Factor (LOF) [46] belong to the commonly-used anomaly detection methods. Among all these compared methods, only SR2CNN can refinedly classify unknown classes.

With the same setting in previous works, we set $\beta = 0.4$ for the SR2CNN to optimize the bounding thresholds in Eq. (27), i.e., perform grid searching Θ_k^* to maximize the

TABLE III
INTRODUCTIONS OF COMPARED METHODS

Setting Method	Requirement of priori data	Goal
S3R [†] (Proposed)	Known classes	Classifying known classes & classifying unknown classes
SR2CNN [‡] [38]	Known classes & unknown classes	Classifying known classes & classifying unknown classes
OpenMax [§] [35]	Known classes	Classifying known classes & rejecting unknown classes
UIOS [§] [36]	Known classes	Classifying known classes & rejecting unknown classes
IF [44]	Known classes	Accepting known classes & rejecting unknown classes
One-Class SVM [45]	Known classes	Accepting known classes & rejecting unknown classes
LOF [46]	Known classes	Accepting known classes & rejecting unknown classes

[†] S3R: <https://github.com/DaftJun/S3R>

[‡] SR2CNN: <https://github.com/YihongDong/SR2CNN-Zero-Shot-Learning-for-Signal-Recognition>

[§] OpenMax: <https://github.com/abhijitbendale/OSDN>

[§] UIOS: <https://github.com/LooKing9218/UIOS>

TABLE IV
SCENARIOS WITH DIFFERENT NUMBERS OF KNOWN AND UNKNOWN CLASSES

Scenario	Splitting setting (A~X refers to class indexes in Table I)	
I	known class (18)	A, B, C, D, F, K, L, M, N, O, P, Q, R, S, T, U, V, X
	unknown class (6)	E, G, H, I, J, W
II	known class (20)	A, B, C, D, F, G, H, K, L, M, N, O, P, Q, R, S, T, U, V, X
	unknown class (4)	E, I, J, W
III	known class (22)	A, B, C, D, F, G, H, I, K, L, M, N, O, P, Q, R, S, T, U, V, W, X
	unknown class (2)	E, J
IV	known class (23)	A, B, C, D, E, F, G, H, I, J, K, L, M, N, P, Q, R, S, T, U, V, W, X
	unknown class (1)	O
V	known class (16)	A, B, C, D, F, K, L, M, N, P, R, S, T, U, V, X
	unknown class (8)	E, G, H, I, J, O, Q, W
VI	known class (14)	A, C, D, F, K, L, M, P, R, S, T, U, V, X
	unknown class (10)	B, E, G, H, I, J, N, O, Q, W

indicator of $0.4 \times \text{TKR} + 0.6 \times \text{TUR}$ in the testing set. Although there are still many embedded hyper-parameters that affect the performance of IF, One-class SVM, and LOF, we tune them to achieve optimal performance on the testing set via the grid search. Besides, these compared detectors require random seeds to perform detection, thus causing significant fluctuations in performance. To be statistically fair, we perform 100 times repetitive experiments and report the best performance of each competitive detector.

D. Accuracy Comparison

Selecting Scenario I in Table IV as a typical scenario, we conduct a detailed indicator comparison with SR2CNN, OpenMax, and UIOS, as shown in Table V. Notice that the unknown accuracy and UP of OpenMax and UIOS cannot be provided due to their inability to refinedly identify unknown signals. Table V shows that all indicators of the proposed S3R method are better than the compared methods. In particular, the average known accuracy is improved by 14.18%, the average unknown accuracy is improved by 18.81%, the TKR improved by 13.61%, and the UP is improved by 17.99%, as compared to the sub-optimal method. Besides, although unknown class Futaba T14SG (W) and known class Futaba T16IZ (T) belong to the same-brand drones, the proposed S3R method successfully achieves 100% unknown accuracy while SR2CNN is only 47.01%, which shows the superiority of the proposed S3R method in identifying same-brand unknown drones. Furthermore, we observe that all methods have poor known accuracy on the DJI Phantom 3 (B), RadioLink AT9S (V), and Futaba T16IZ (T), because they do not transmit VTS.

Therefore, it is more challenging to identify drones only by the FCS when the VTS are cut off or not transmitted.

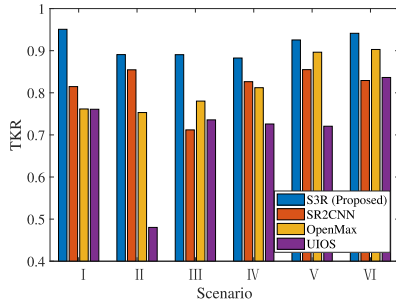
In Fig. 6, we demonstrate the generalizability of the proposed method facing the scenarios with different numbers of known and unknown classes. Across the six preset scenarios in Table IV, the proposed S3R method has achieved significant advantages in most indicators. Specifically, S3R can keep the TUR greater than 95% in all scenarios, which indicates that S3R can more effectively reject unknown signals. In addition, the indicators of the compared methods fluctuate greatly when the scenario changes, which indicates that the generalizability of the compared methods is weak. Furthermore, among all the preset scenarios, Scenario VI has the smallest closed set and the largest open set, which makes the recognition of Scenario VI more challenging. Nevertheless, the generalizability of the proposed S3R method is superior, which is reflected in the significant leading indicators of the TUR and the KP in Scenario VI.

E. Generalizability Comparison

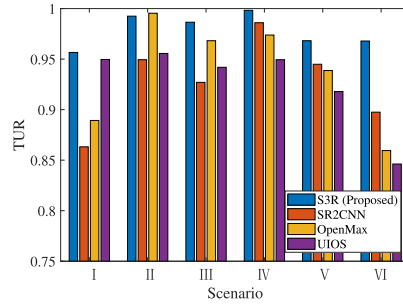
The generalizability also reflects in the sensitivity of the proposed method to the scenarios where different unknown classes are combined. To this end, three scenarios are provided in Table VI by adjusting Scenario I, namely, Scenario I-A, Scenario I-B, and Scenario I-C. In particular, unknown classes in these three scenarios are restricted to be completely different from each other. Therefore, these three scenarios can be used to evaluate the sensitivity in dealing with unfamiliar scenarios. As shown in Fig. 7, we can summarize the following findings. First of all, in terms of the KP, the task of identifying

TABLE V
DETAILED COMPARISONS IN SCENARIO I

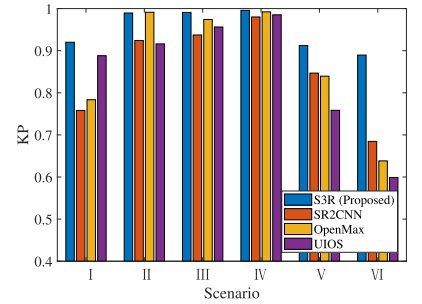
Method		S3R (ours)	SR2CNN	OpenMax	UIOS
Indicator					
Known accuracy	Background (A)	0.9683	0.8333	0.0317	0.8413
	DJI Phantom 3 (B)	0.9124	0.8613	0.6131	0.9708
	DJI Phantom 4 Pro (C)	0.9655	0.7034	0.3586	0.8345
	DJI MATRICE 200 (D)	0.9728	0.9116	1.000	0.0612
	DJI Air 2S (F)	0.9740	0.9221	0.9221	1.0000
	DJI Mavic 3 (K)	0.9381	0.7938	0.7835	0.9381
	DJI MATRICE 300 (L)	0.9474	0.7895	0.2281	0.9298
	DJI Phantom 4 Pro RTK (M)	0.9580	0.9244	0.8655	1.0000
	DJI MATRICE 30T (N)	0.9706	0.7500	0.7941	0.5000
	DJI AVATA (O)	0.9029	0.8835	0.9029	0.0000
	DJI DIY (P)	0.9062	0.8750	0.8984	0.9141
	DJI MATRICE 600 Pro (Q)	0.9524	0.7524	0.9619	0.9810
	VBar (R)	0.9748	0.6723	0.8235	0.9076
	FrSky X20 (S)	1.0000	0.7477	0.9813	1.0000
	Futaba T16IZ (T)	0.8846	0.7385	0.8615	0.0000
	Taranis Plus (U)	0.9754	0.8607	0.8033	1.0000
	RadioLink AT9S (V)	0.9417	0.7087	0.8350	1.0000
	Skydroid (X)	0.9851	0.8507	0.9851	1.0000
Unknown accuracy	DJI MATRICE 100 (E)	0.9020	0.8136	-	-
	DJI Mini 3 Pro (G)	0.5000	0.5042	-	-
	DJI Inspire 2 (H)	0.8680	0.4299	-	-
	DJI Mavic Pro (I)	1.0000	0.9770	-	-
	DJI Mini 2 (J)	0.9879	0.9341	-	-
	Futaba T14SG (W)	1.0000	0.4701	-	-
Average known accuracy		0.9517	0.8099	0.7583	0.7710
Average unknown accuracy		0.8763	0.6882	-	-
TKR		0.9508	0.8147	0.7616	0.7606
TUR		0.9566	0.8632	0.8893	0.9496
KP		0.9201	0.7580	0.7835	0.8881
UP		0.8763	0.6964	-	-



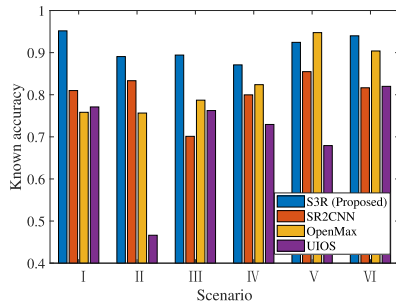
(a) TKR



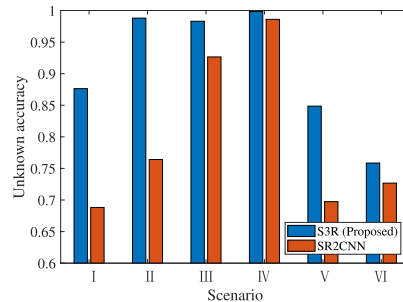
(b) TUR



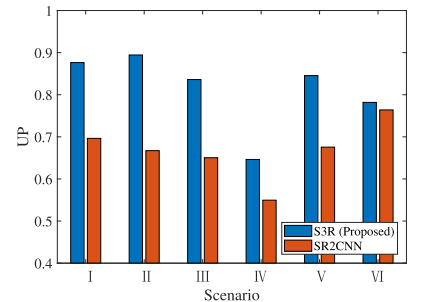
(c) KP



(d) Known accuracy



(e) Unknown accuracy



(f) UP

Fig. 6. Indicator comparison across multi-scenarios.

Scenario I-C is the most challenging. Nonetheless, the KP decrease of the proposed method is the least significant as the task difficulty increases. Second, accepting known classes and rejecting unknown classes are often trade-offs for OSR

methods. The OpenMax can maintain a superior ability to accept known classes when facing different scenarios, which is slightly better than the proposed method. Nevertheless, considering the ability to reject unknown classes, the proposed

TABLE VI
SCENARIOS WITH DIFFERENT COMBINATIONS OF KNOWN AND UNKNOWN CLASSES

Scenario	Splitting setting (A~X refers to class indexes in Table I)	
I-A	known class	B, C, D, E, F, G, H, I, J, L, P, Q, R, S, T, V, W, X
	unknown class	A, K, M, N, O, U
I-B	known class	A, C, D, F, G, H, I, J, K, M, N, O, P, S, T, U, V, X
	unknown class	B, E, L, Q, R, W
I-C	known class	A, B, D, E, G, H, I, K, L, M, N, O, Q, R, U, V, W, X
	unknown class	C, F, J, P, S, T

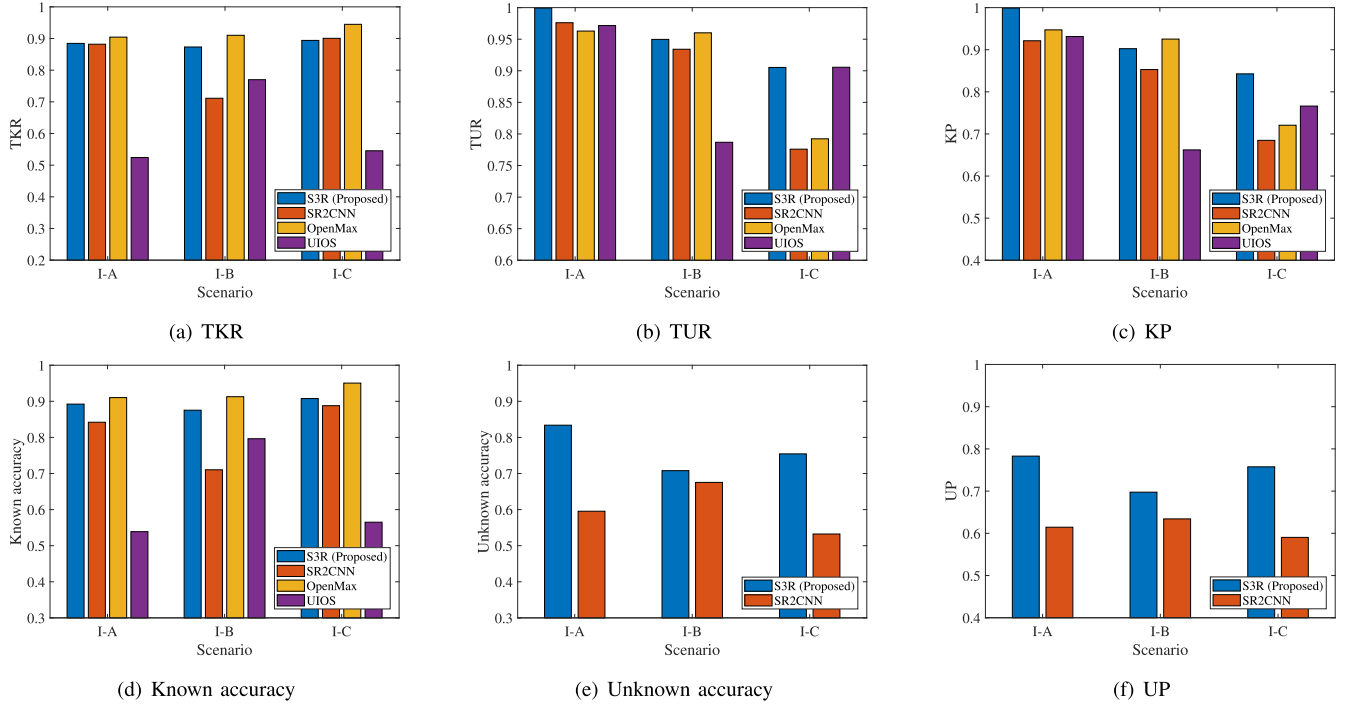


Fig. 7. Sensitivity to the different combinations of known and unknown classes.

TABLE VII
VALIDITY VERIFICATION OF PROPOSED TE AND PE

Indicator \ Extractor	TE & PE	Only TE	Only PE
Known accuracy	0.9517	0.9171	0.8591
Unknown accuracy	0.8763	0.7180	0.4379
TKR	0.9508	0.9149	0.8559
TUR	0.9566	0.9118	0.9258
KP	0.9201	0.8451	0.8580
UP	0.8763	0.7489	0.4191

S3R method performs best, in terms of the higher TUR and the smaller fluctuations as the scenario changes. Third, the proposed S3R method significantly outperforms SR2CNN in all the three scenarios and achieves an average gain of the UP over 10%.

F. Ablation Study

First, we verify the importance of the designed TE and PE in S3R, as shown in Table VII. We can see that S3R equipped with both TE and PE performs better than S3R equipped with only TE or PE. Specifically, S3R with TE can only achieve an unknown accuracy of 71.80%, but can achieve an improved unknown accuracy of 87.63% with the assistance of PE. On the other hand, the performance of PE is much inferior

to TE, which shows that texture features are more important than position features in identifying drone RF signals via TFS awareness. With the above-mentioned results, it is concluded that both the proposed TE and PE play a fundamental role in the feature extraction of the TFS.

Then, we study the indicators of the proposed method versus different hyper-parameter settings. As shown in Fig. 8(a), we can infer that the choice of the dimension of semantics is important. Specifically, a larger D can bring a better TUR and KP, while making the TKR and the UP decreased. Considering the all indicators, it is a relatively suitable choice to set D in the range of 32 ~ 128. As shown in Fig. 8(b), ensuring the signal length is critical to improve all indicators, since the longer signals can carry more signal characteristics. As shown in Fig. 8(c), increasing the number of training samples can significantly improve the TKR and the UP, but slightly degrade the TUR and the KP. As shown in Fig. 8(d), we can see that a larger ϑ can improve the TUR and the KP, while making the TKR and the UP worse when ϑ is set to more than 8. Considering the above results, we set ϑ as 8 and D as 128 in this paper.

Next, we verify the advantages of the proposed outlier analysis-based classifier, compared to commonly-used anomaly detectors, i.e., IF, One-class SVM, and LOF. It should

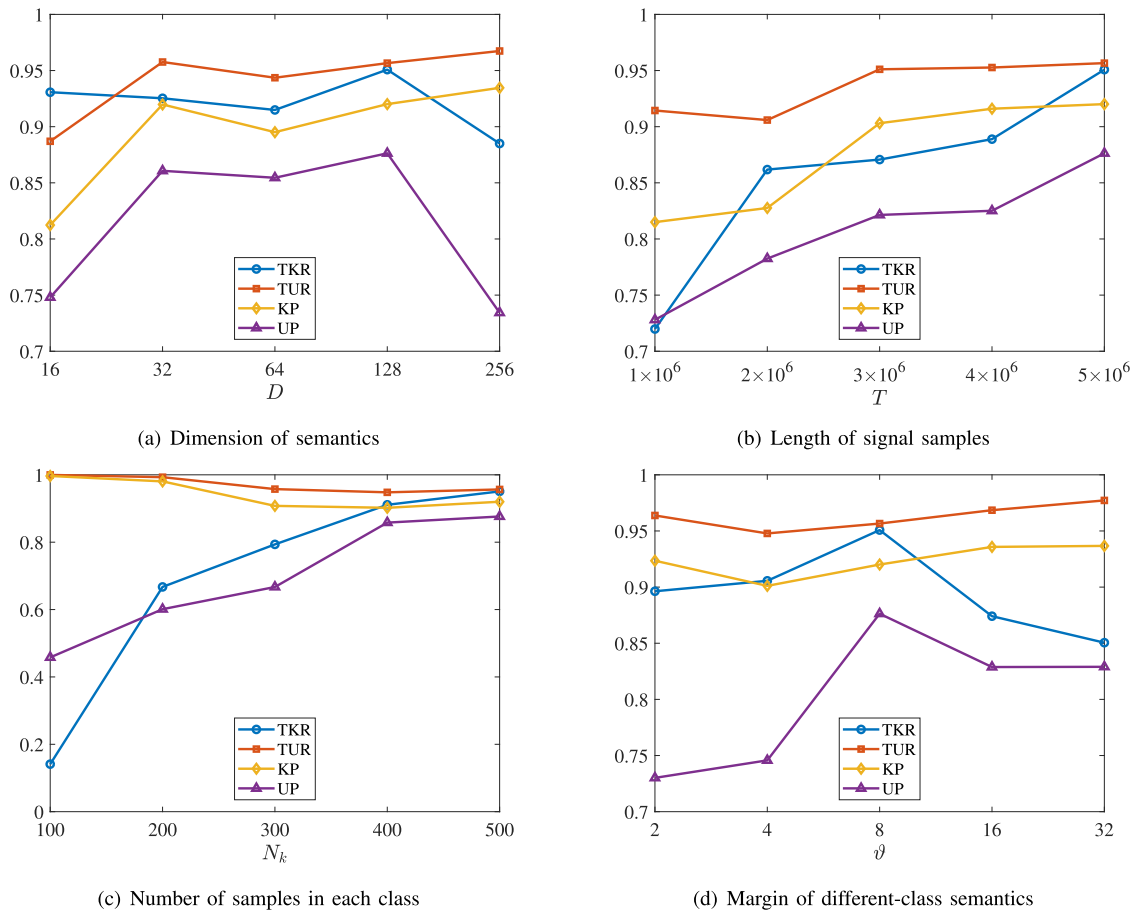


Fig. 8. Indicator curves with different settings of hype-parameters.

TABLE VIII
VALIDITY VERIFICATION OF PROPOSED OUTLIER
ANALYSIS-BASED CLASSIFIER

Method \ Indicator	TKR	TUR	Average
Proposed	0.9508	0.9566	0.9537
IF	0.8791	0.9247	0.9019
One-class SVM	0.8864	0.9783	0.9324
LOF	0.8555	0.9775	0.9165

be pointed out that since the compared detectors do not have their own semantic representations, the proposed S3R method shares its semantics with them during comparison. In this way, the performance gaps are entirely determined by the semantic classifiers themselves. As shown in Table VIII, the proposed S3R method has the best TKR, particularly with a gain of 6.44%. In terms of TUR, the proposed S3R method is only 2.17% inferior to the One-class SVM. Obviously, the compared anomaly detectors sacrifice more TKR in exchange for TUR's leads. But in OSR, TKR and TUR are equally important. Therefore, comparing the average of TKR and TUR, the proposed S3R method is the best with an average gain of 2.13%. As all results are obtained statistically with repeated experiments, the performance lead of the proposed S3R method is reliable.

Finally, it is verified that the proposed augmented semantics can effectively improve the recognition of unknown classes.

TABLE IX
VALIDITY VERIFICATION OF AUGMENTED SEMANTICS

Indicator \ Semantic	Augmented	Raw
Unknown accuracy	0.9517	0.9171
UP	0.8763	0.7180

As shown in Table IX, by employing the augmented semantics, S3R can improve the unknown accuracy and the UP by 3.46% and 15.83%, respectively. This demonstrates that the augmented semantics have a stronger ability to represent unknown signals than the raw semantics, and thus can perform well in the refined recognition of the open set because the basic features are fused.

G. Visualization Interpretation

The performance difference of the compared methods can be reflected in the structural distribution of the testing samples in the semantic space. First of all, different from the semantic space construction of the other three methods, the semantic vectors of UIOS are distributed in a long strip shape because of the SoftPlus activation layer at the end of the model. Second, all methods can effectively identify DJI Mini 3 Pro as an unknown class, but they consider it to be two unknown classes, corresponding to the yellow clusters in the semantic space. It is reasonable because newly-released drone products often have two different signal modulation modes, due to the need for a

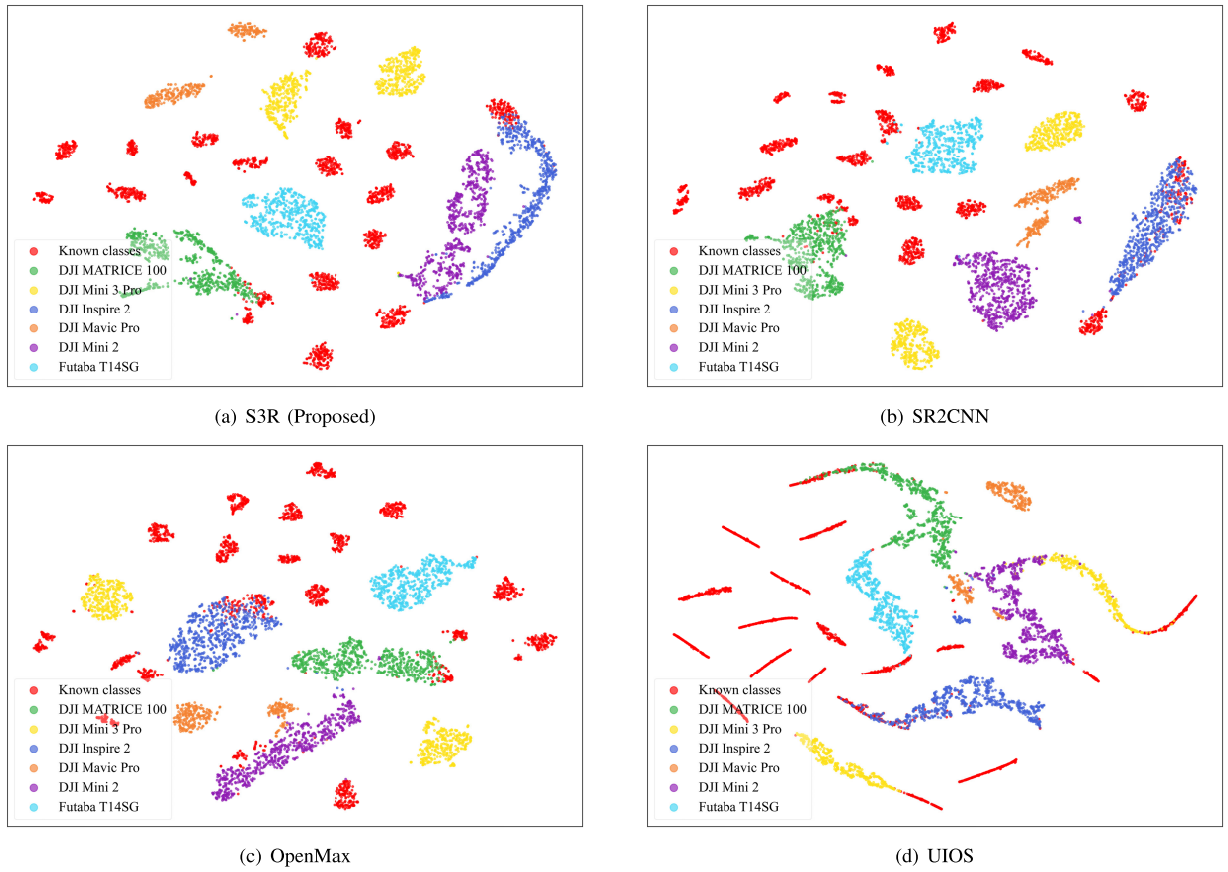


Fig. 9. Visualization of signal semantic spaces regarding Scenario I.

better anti-interference performance. According to our investigation on the dataset after the experiment, there are indeed two signal modulation modes present in the communication of DJI Mini 3 Pro. However, since the priori labelling assigns these two modes to the same class, both S3R and SR2CNN achieve only an about 50% unknown accuracy for DJI Mini 3 Pro.

Another important finding is that the unknown DJI Inspire 2 (blue instances) and the known DJI MATRICE 200 (red instances) are highly overlapping in the semantic space of both SR2CNN and OpenMax. It is because these two products have similar texture features in the spectrum, i.e., the signal duration and bandwidth of FCS and VTS. Nevertheless, thanks to the effectiveness of PE, fewer testing samples overlap in the semantic space of S3R. Consequently, S3R performs much better on classifying the unknown DJI Inspire 2 than SR2CNN, as indicated in Table V.

The visual reflection of Fig. 9(a) and Fig. 9(b) indicates that SR2CNN exhibits a stronger sparsity for semantic clustering of unknown classes in the raw semantic space. Nevertheless, S3R still achieves leading performance on the UP after the proposed semantic augmentation, as shown in Table V. This highlights the necessity of semantic augmentation for the classification of unknown classes.

VII. CONCLUSION

In this paper, we proposed an S3R method for RF-based drone recognition in the open set scenario. The formulated extractors extract the texture and position features from the signal spectra to construct distinguishable signal semantics.

The bounding thresholds for the signal semantic classification are self-adaptively obtained by analyzing outliers of each known class in the closed set. By fusing the basic features from the intermediate layers of the extractor, the signal semantics are augmented to be clustered for the refined unknown recognition. The recognition accuracy and generalizability of the proposed method are verified through extensive experiments compared with existing methods.

APPENDIX A

THE ILLUSTRATION OF TRANSLATION INVARIANCE IN TE

The operation of pooling is popular in modern NN, as it can expand the receptive field and reduce the number of learnable parameters. Assuming that the size of the feature maps is $\mathbb{R}^{4 \times 4}$, two possible feature maps corresponding to a pair of same-brand drones (named Drone-A and Drone-B, respectively) can be denoted as

$$\mathbf{Z}_A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \mathbf{Z}_B = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad (37)$$

where 1 represents that there exists the signal of the drone, and 0 represents that there is no signal. The operation of MaxPooling can be denoted as

$$\text{MaxPool}(\mathbf{Z}_A) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \text{MaxPool}(\mathbf{Z}_B) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (38)$$

The operation of the GlobalAvgPooling can be denoted as

$$\begin{aligned}\text{GlobalAvgPool}(\mathbf{Z}_A) &= [0.125], \\ \text{GlobalAvgPool}(\mathbf{Z}_B) &= [0.125].\end{aligned}\quad (39)$$

It can be seen that the generated feature maps after pooling are the same, although the position features of Drone-A and Drone-B are different. This indicates that the position features are inevitably ignored by TE due to its translation invariance.

APPENDIX B

THE NECESSITY OF POSITION ENCODING FOR PE

According to Eq. (11), there is a position encoding denoted as $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_n, \dots, \mathbf{p}_N]$, where $\mathbf{p}_n \in \mathbb{R}^W$. Assuming that two FCS of the drone are respectively located at the positions with the index n and $n + \Delta n$, where Δn denotes the interval of the relative position, the result of the feature mapping in the SA module can be expressed as

$$\begin{aligned}[\mathbf{p}_n]^T \mathbf{p}_{n+\Delta n} &= \sum_{w=1}^{\frac{W}{2}} [\cos \omega_w n \cdot \cos \omega_w (n + \Delta n) \\ &\quad + \sin \omega_w n \cdot \sin \omega_w (n + \Delta n)] \\ &= \sum_{w=1}^{\frac{W}{2}} \cos \omega_w \Delta n,\end{aligned}\quad (40)$$

From Eq. (40), it is clear that the features of the relative position, i.e., Δn , can be easily learned by PE.

REFERENCES

- [1] S. Hu, W. Ni, X. Wang, A. Jamalipour, and D. Ta, "Joint optimization of trajectory, propulsion, and thrust powers for covert UAV-on-UAV video tracking and surveillance," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1959–1972, 2021.
- [2] X. Shi, C. Yang, W. Xie, C. Liang, Z. Shi, and J. Chen, "Anti-drone system with multiple surveillance technologies: Architecture, implementation, and challenges," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 68–74, Apr. 2018.
- [3] P. Chen, Z. Chen, B. Zheng, and X. Wang, "Efficient DOA estimation method for reconfigurable intelligent surfaces aided UAV swarm," *IEEE Trans. Signal Process.*, vol. 70, pp. 743–755, 2022.
- [4] P. Araujo, J. Fontinele, and L. Oliveira, "Multi-perspective object detection for remote criminal analysis using drones," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1283–1286, Jul. 2020.
- [5] J. Du, X. Luo, L. Jin, and F. Gao, "Robust tensor-based algorithm for UAV-assisted IoT communication systems via nested PARAFAC analysis," *IEEE Trans. Signal Process.*, vol. 70, pp. 5117–5132, 2022.
- [6] T. Z. H. Ernest, A. S. Madhukumar, R. P. Sirigina, and A. K. Krishna, "NOMA-aided UAV communications over correlated Rician shadowed fading channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 3103–3116, 2020.
- [7] C. Shen, T.-H. Chang, J. Gong, Y. Zeng, and R. Zhang, "Multi-UAV interference coordination via joint trajectory and power control," *IEEE Trans. Signal Process.*, vol. 68, pp. 843–858, 2020.
- [8] W. Tang, H. Zhang, Y. He, and M. Zhou, "Performance analysis of multi-antenna UAV networks with 3D interference coordination," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 5145–5161, Jul. 2022.
- [9] Y. Xie, P. Jiang, Y. Gu, and X. Xiao, "Dual-source detection and identification system based on UAV radio frequency signal," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2021.
- [10] M. F. Al-Sa'd, A. Al-Ali, A. Mohamed, T. Khatlab, 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, Nov. 2019.
- [11] C. Xu, B. Chen, Y. Liu, F. He, and H. Song, "RF fingerprint measurement for detecting multiple amateur drones based on STFT and feature reduction," in *Proc. Integr. Commun. Navigat. Surveill. Conf. (ICNS)*, Sep. 2020, p. 4G1.
- [12] R. Kılıç, N. Kumbasar, E. A. Oral, and I. Y. Ozbek, "Drone classification using RF signal based spectral features," *Eng. Sci. Technol., Int. J.*, vol. 18, Apr. 2022, Art. no. 101028.
- [13] O. O. Medaiyese, M. Ezuma, A. P. Lauf, and I. Guvenc, "Wavelet transform analytics for RF-based UAV detection and identification system using machine learning," *Pervas. Mobile Comput.*, vol. 82, Jun. 2022, Art. no. 101569.
- [14] C. Xu, F. He, B. Chen, Y. Jiang, and H. Song, "Adaptive RF fingerprint decomposition in micro UAV detection based on machine learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 7968–7972.
- [15] A. Shoufan, H. M. Al-Angari, M. F. A. Sheikh, and E. Damiani, "Drone pilot identification by classifying radio-control signals," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 10, pp. 2439–2447, Oct. 2018.
- [16] Q. Wang, L. Wang, L. Yu, J. Wang, and X. Zhang, "An ID-based robust identification approach toward multipoint noncooperative drones," *IEEE Sensors J.*, vol. 23, no. 9, pp. 10179–10192, May 2023.
- [17] S. Rajendran, W. Meert, V. Lenders, and S. Pollin, "Unsupervised wireless spectrum anomaly detection with interpretable features," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 3, pp. 637–647, Sep. 2019.
- [18] Z. Cai, Y. Wang, Q. Jian, G. Gui, and J. Sha, "Toward intelligent lightweight and efficient UAV identification with RF fingerprinting," *IEEE Internet Things J.*, early access, Apr. 30, 2024, doi: 10.1109/IJOT.2024.3395466.
- [19] A. A. Fanid, M. Dabaghchian, N. Wang, P. Wang, L. Zhao, and K. Zeng, "Machine learning-based delay-aware UAV detection and operation mode identification over encrypted Wi-Fi traffic," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2346–2360, 2020.
- [20] J. Deng, X. Ji, B. Wang, B. Wang, and W. Xu, "Dr. Defender: Proactive detection of autopilot drones based on CSI," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 194–206, 2024.
- [21] L. Lin, N. Yu, Y. Wang, and Z. Shi, "5G spectrum learning-based passive UAV detection in urban scenario," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2023, pp. 1–5.
- [22] C. Geng, S. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3614–3631, Oct. 2021.
- [23] X. Zhang et al., "Enhanced few-shot malware traffic classification via integrating knowledge transfer with neural architecture search," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 5245–5256, 2024.
- [24] H. Fu, S. Abeywickrama, L. Zhang, and C. Yuen, "Low-complexity portable passive drone surveillance via SDR-based signal processing," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 112–118, Apr. 2018.
- [25] M. Ezuma, C. K. Anjinappa, M. Funderburk, and I. Guvenc, "Radar cross section based statistical recognition of UAVs at microwave frequencies," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 1, pp. 27–46, Feb. 2022.
- [26] C. Wang, J. Tian, J. Cao, and X. Wang, "Deep learning-based UAV detection in pulse-Doppler radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5105612.
- [27] B. K. Kim, H.-S. Kang, and S.-O. Park, "Drone classification using convolutional neural networks with merged Doppler images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 38–42, Jan. 2017.
- [28] Y. Sun, S. Abeywickrama, L. Jayasinghe, C. Yuen, J. Chen, and M. Zhang, "Micro-Doppler signature-based detection, classification, and localization of small UAV with long short-term memory neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6285–6300, Aug. 2021.
- [29] B. Torvik, K. E. Olsen, and H. Griffiths, "Classification of birds and UAVs based on radar polarimetry," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1305–1309, Sep. 2016.
- [30] M. W. Ashraf, W. Sultani, and M. Shah, "Dogfight: Detecting drones from drones videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7063–7072.
- [31] H. Wang, X. Wang, C. Zhou, W. Meng, and Z. Shi, "Low in resolution, high in precision: UAV detection with super-resolution and motion information extraction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.
- [32] Z. Shi, X. Chang, C. Yang, Z. Wu, and J. Wu, "An acoustic-based surveillance system for amateur drones detection and localization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2731–2739, Mar. 2020.

- [33] J. Fang, Y. Li, P. N. Ji, and T. Wang, "Drone detection and localization using enhanced fiber-optic acoustic sensor and distributed acoustic sensing technology," *J. Lightw. Technol.*, vol. 41, no. 3, pp. 822–831, Feb. 6, 2023.
- [34] 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.
- [35] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1563–1572.
- [36] M. Wang et al., "Uncertainty-inspired open set learning for retinal anomaly identification," *Nature Commun.*, vol. 14, no. 1, p. 6757, Oct. 2023.
- [37] T. Li et al., "The importance of expert knowledge for automatic modulation open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13730–13748, Nov. 2023.
- [38] Y. Dong, X. Jiang, H. Zhou, Y. Lin, and Q. Shi, "SR2CNN: Zero-shot learning for signal recognition," *IEEE Trans. Signal Process.*, vol. 69, pp. 2316–2329, 2021.
- [39] H. Han, W. Li, Z. Feng, G. Fang, Y. Xu, and Y. Xu, "Proceed from known to unknown: Jamming pattern recognition under open-set setting," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 693–697, Apr. 2022.
- [40] 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, 2023.
- [41] P. Deng, S. Hong, J. Qi, L. Wang, and H. Sun, "A lightweight transformer-based approach of specific emitter identification for the automatic identification system," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 2303–2317, 2023.
- [42] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, "Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2203–2215, Sep. 2018.
- [43] F. Nie, Z. Li, R. Wang, and X. Li, "An effective and efficient algorithm for k-means clustering with new formulation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3433–3443, Apr. 2023.
- [44] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Min.*, Apr. 2008, pp. 413–422.
- [45] Y. Chen, X. S. Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," in *Proc. Int. Conf. Image Process.*, 2001, pp. 34–37.
- [46] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," *ACM SIGMOD Rec.*, vol. 29, no. 2, pp. 93–104, May 2000.



Ningning Yu (Student Member, IEEE) received the B.S. and M.S. degrees in information engineering from Zhejiang University of Technology, Hangzhou, China, in 2018 and 2021, respectively. He is currently pursuing the Ph.D. degree in electronic science and technology with Zhejiang University, Hangzhou. His research interests include anti-drone technology, spectrum recognition, and machine learning.



Jiajun Wu received the B.S. degree in electronic science and technology from Zhejiang University, Hangzhou, China, in June 2024, where he is currently pursuing the Ph.D. degree in electronic science and technology. His research interests include anti-drone technology, signal processing, and pattern recognition.



Chengwei Zhou (Senior Member, IEEE) received the Ph.D. degree in electronic science and technology from Zhejiang University, Hangzhou, China, in June 2018.

He was a Visiting Researcher with the University of Technology Sydney, Sydney, NSW, Australia, from April 2017 to October 2017. He was a Post-Doctoral Research Fellow with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, from 2018 to 2020. Since June 2020, he has been with the College of Information Science and Electronic Engineering, Zhejiang University, where he is currently a tenure-track Professor. His current research interests are in the areas of array signal processing, direction-of-arrival estimation, and adaptive beamforming. He was a recipient of the 2021 IEEE Signal Processing Society Young Author Best Paper Award, the IEEE ICASSP 2023 Top 3% Paper Recognition, the 2019 IET Communications Premium Award, and the ISAP 2020 Best Paper Award. He was a Committee Member of several IEEE international conferences, including the 11th IEEE Sensor Array and Multichannel Signal Processing Workshop (IEEE SAM 2020) and the 13th IEEE/CIC International Conference on Communications in China (IEEE/CIC ICC 2024). He serves as an Associate Editor for *International Journal of Communication Systems* and *Franklin Open* and an Editorial Board Member of *Journal of Signal Processing* (in Chinese).



Zhiguo Shi (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Zhejiang University, Hangzhou, China, in 2001 and 2006, respectively.

Since 2006, he has been a Faculty Member with the College of Information Science and Electronic Engineering, Zhejiang University, where he is currently a Full Professor. From 2011 to 2013, he visited the Broadband Communications Research Group, University of Waterloo, Waterloo, ON, Canada. His research interests include array signal processing, localization, and the Internet of Things. He is an Elected Member of the Sensor Array and Multichannel (SAM) Technical Committee IEEE Signal Processing Society. He was a recipient of the 2019 IET Communications Premium Award, and co-authored a paper that received the 2021 IEEE Signal Processing Society Young Author Best Paper Award. He was also a recipient of the Best Paper Award from ISAP 2020, IEEE GLOBECOM 2019, IEEE WCNC 2017, IEEE/CIC ICC 2013, and IEEE WCNC 2013. He was the General Co-Chair of IEEE SAM 2020. He served as an Editor for IEEE NETWORK. He is currently serving as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and *Journal of the Franklin Institute*.



Jiming Chen (Fellow, IEEE) received the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2005. He is currently a Professor with the Department of Control Science and Engineering, and the Vice Dean with the Faculty of Information Technology, Zhejiang University. His research interests include the IoT, networked control, and wireless networks. He is a fellow of CAA. He was a recipient of the Seventh IEEE ComSoc Asia/Pacific Outstanding Paper Award, the JSPS Invitation Fellowship, and the IEEE ComSoc AP Outstanding Young Researcher Award. He was the General Co-Chair of IEEE RTCSA'19, IEEE Datacom'19, and IEEE PST'20. He serves on the editorial boards of multiple IEEE TRANSACTIONS. He is an IEEE VTS Distinguished Lecturer.