Radio Frequency Fingerprint Collaborative Intelligent Identification Using Incremental Learning

Mingqian Liu, *Member, IEEE,* Jiakun Wang, Nan Zhao, *Senior Member, IEEE,* Yunfei Chen, *Senior Member, IEEE,* Hao Song, and F. Richard Yu, *Fellow, IEEE*

***Abstract*—For distributed sensor systems using neural networks, each sub-network has a different electromagnetic environment, and these recognition accuracy is also different. In this paper, we propose a distributed sensor system using incremental learning to solve the problem of radio frequency fingerprint identification. First, the intelligent representation of the received signal is linearly fused into a four-channel image. Then, convolutional neural network is trained by using the existing data to obtain the preliminary model of the network, and decision fusion is used to solve the problem in the distributed system. Finally, using new data, instead of retraining the model, we employ incremental learning by fine-tuning the preliminary model. The proposed method can significantly reduce the training time and is adaptive to streaming data. Extensive experiments show that the proposed method is computationally efficient, and also has satisfactory recognition accuracy, especially at low signal-to-noise ratio (SNR) regime.**

***Index Terms*—Collaborative identification, convolutional neural network, incremental learning, radio frequency fingerprint identification**

I. Introduction

RADIO frequency fingerprint identification (RF-FI) refers to the technology that only uses the external signal to extract the information about the target identity (known as "emitter fingerprint"), and searches the fingerprint in the feature library, to determine the specific emitter transmitting a given signal [1]–[6]. Due to its unique role in identifying specific originators, RF-FI technology has attracted extensive attention in spectrum management, network security, cognitive radio and electronic countermeasures. The traditional method mainly uses the feature analysis (of the data and signal) and feature extraction based on which the feature template is constructed and the recognition library is generated to recognize the specific signal. As the interference in electromagnetic environments increases, accurate individual recognition has become extremely difficult and expensive. In the distributed sensor networks, due to the influence of different propagation and transmission environment, even if the transmitter sends the same signal, different signals will be observed on each receiving sensor. If the complete sensor observation data can be transmitted to the master node, the difference information between signals can be fully reflected, and the global optimization of RF-FI can be realized. Therefore, RF-FI based on distributed networks has strong application value [7]–[10].

S. Guo, C. Song et al. proposed an empirical mode decomposition (EMD) method to extract features, but this method has modal aliasing [11], [12]. Y. Liu et al. introduced a feature extraction method based on Hilbert Huang transform (HHT), which combines EMD algorithm with Hilbert transform to extract instantaneous frequency and amplitude. However, it has endpoint effect [13]. The authors in [14]–[17] put forward a method of feature extraction based on bi-spectrum transformation (BST). This method has high feature dimension. Y. Li, G. Lopez-Risueo et al. proposed a method of feature extraction based on STFT, which extracts the time-frequency distribution for fingerprint recognition , but for this method is not applicable to nonlinear signal processing [18], [19]. K. Merchant et al. introduced a method of feature extraction by dividing the signal into two I/Q channels. However, this method is sensitive to noise. Additionally, the length of the signal is too long, which is not conducive to the improvement of neural network training speed [20]. A feature extraction method based on differential constellation (DCTF) was proposed by L. Peng et al. This method has a high requires good time synchronization [21], [22], which may not be practical in reality. In [23], the transmitter internal noise and device modeling were used for individual identification based on the unintentional modulation characteristics. Nevertheless, this method can only model local devices and links, making it not suitable in large-scale individual identification. To cope with the aforementioned issues, we will propose a recognition scheme that combines HHT, short-time Fourier transform (STFT), ambiguity function, and bi-spectrum transform for signal processing, which could fully mine the fingerprint information difference between signals.

Moreover, traditional classifiers mostly use shallow learner,

This work was supported by the National Natural Science Foundation of China under Grant 62071364, in part by the Aeronautical Science Foundation of China under Grant 2020Z073081001, in part by the Fundamental Research Funds for the Central Universities under Grant JB210104, and in part by the 111 Project under Grant B08038. This paper have been presented in part at the 2021 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS 2021): Mingqian Liu, Jiakun Wang and Qian Cheng, Incremental Learning Based Radio Frequency Fingerprint Identification Using Intelligent Representation. *(Corresponding author: Jiakun Wang.)*

M. Liu and K. Wang are with the State Key Laboratory of Integrated Service Networks, Xidian University, Shaanxi, Xi'an 710071, China (e-mail: mqliu@mail.xidian.edu.cn; wangjiakun@stu.xidian.edu.cn).

N. Zhao is with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: zhaonan@dlut.edu.cn).

Y. Chen is with the School of Engineering, University of Warwick, Coventry, West Midlands United Kingdom of Great Britain and Northern Ireland CV4 7AL (e-mail: yunfei.chen@warwick.ac.uk).

H. Song is with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24060, USA (e-mail: haosong@vt.edu).

F. R. Yu is with the Department of System and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: richard.yu@carleton.ca).

such as Support Vector Machine (SVM) and K-means mean clustering [24]. They heavily depend on feature extraction. If the extracted features are not able to show the difference between signals well, the recognition accuracy will be greatly reduced. In recent years, relevant studies have shown that deep learning networks have better recognition accuracy than shallow learners, because its unique structure can mine deep features of data. As a result, deep learning networks have been widely used in image recognition and wireless signal classification [25]–[29].

In practice, the data that need to be trained may not be available at the same time when new data comes. If all the data are combined for retraining, good accuracy may be obtained. However, the previous training wasted. Moreover, it incur large overhead and a long training time is needed by storing previous data. Recently, incremental learning is introduced to solve this issue. When part of the data are available, they will be trained. To meet the needs of industry, it has gradually become the current mainstream training method [30]. The authors of [31]–[33] proposed an incremental training scheme using part of the old data. When building the old sample set, the data with great influence on the weight was selected and the loss function was also modified, which can not only shorten the training time but also achieve satisfactory recognition accuracy. Z. Li, S. Dang et al. proposed an incremental training method that does not use the old data at all [34], [35]. The training of this method is faster, and the recognition accuracy is not as good as the former, but is satisfactory.

In this paper, we propose a distributed incremental learning method based on convolutional neural network to train the network and realize RF-FI identification. Distributed system can overcome the shortcomings of low accuracy of RF-FI identification and poor adaptability to the environment in complex electromagnetic environment. The main contributions of this paper are as follows:

- The signal is characterized by Hilbert-Huang transform, bi-spectrum transform, short-time Fourier transform, ambiguity function. Then these features are linearly fused to fully display the various features of the signal for distributed sensor system.
- Deep learning framework is adopted, in which the data is trained by using convolutional neural network. It avoids the problem that the shallow network classifier has with poor classification results due to the insufficient feature information.
- Convolution neural network adopts incremental learning. First, the original data are input to the network for training, and the model parameters are saved. Then, when there are new classes of data input, incremental training is carried out by using or not using the old data, which not only shortens the training time but also has satisfactory recognition accuracy.
- Using incremental learning to solve the current complex electromagnetic environment, streaming data storage and training problems, and reduce the space required to store data and shorten the training time, while the recognition accuracy is still satisfactory.

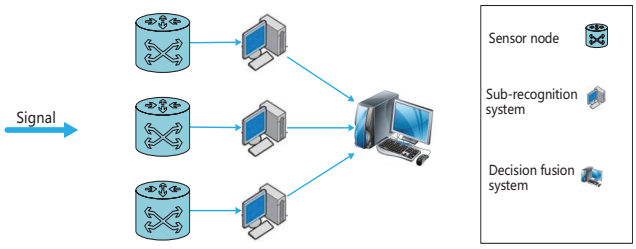

Fig. 1. Radio frequency fingerprint collaborative intelligent identification system model.

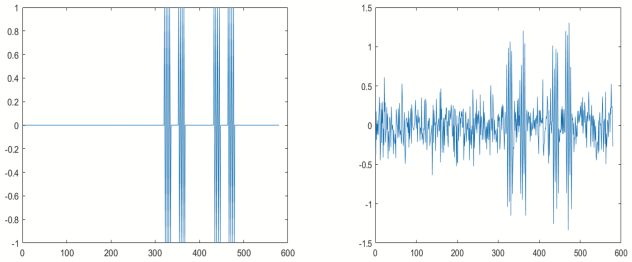

Fig. 2. ADS-B noise-free signal (left) and noisy signal (right).

The remainder of this paper is organized as follows. In Section II, the system is introduced. In Section III, the intelligence representation of signals with fingerprint information is shown. The incremental convolutional neural networks are discussed in Section IV. Simulation studies are given in Section V. Finally, Section VI concludes the whole paper.

II. System Model

In this paper, we consider a distributed learning networks architecture, and the radio frequency fingerprint collaborative intelligent identification system model is shown in Fig. 1. Receiving signals at each sensor and using incremental learning to identify RF-FI as a sub-recognition system. Finally, the output value of each sensor is fused through a decision fusion system. In the system, the received signal can be written as:

$$X(t) = HS(t) + N(t), \tag{1}$$

where $S(t)$ is the automatic dependent surveillance CE broadcast (ADS-B) signal, H stands for the transmission channel and $N(t)$ represents the additive Gaussian noise.

As a common radiation source, ADS-B signal is often used to verify the RF-FI identification technology of radiation source. Therefore, this paper adopts ADS-B signals as input [22], [23]. ADS-B is mainly based on the Mode S 1090 ES and UAT data link transmission using pulse position (PPM) coding. The ADS-B 1090 ES signal has a total of four pulses. The duration of each pulse is 0.5± 0.05 ms ADS-B message data pulse of the first bit first appeared in a relatively leading pulse position ms 8.0 each length of ADS-B message data fields are in 112 bits, the signal is processed by encoding the message through PPM.

Fig. 2 shows an example of ADS-B signal with and without noise. For the received ADS-B signal, we first use four methods of feature extraction as HHT, STFT, AF and BST. Then, the features are combined linearly to obtain a new feature graph. The new feature map is input into the neural network for classification and recognition to obtain the training model and the incremental learning is used for incremental training. Finally, decision fusion is used to solve the problem of recognition accuracy reduction in distributed sensor system.

III. Intelligent Representation of Fingerprint

A. Hilbert Huang Transform

Hilbert transform is a very important method in signal processing. The signals processed by Hilbert transform are generally linear steady-state, but most signals to be processed do not meet this requirement. The empirical mode decomposition (EMD) algorithm can make the time-domain signals linear steady-state. Huang proposed the HHT algorithm which combines EMD and Hilbert spectrum.

The purpose of EMD decomposition is to decompose a signal $f(t)$ into n intrinsic mode functions (IMFs) and a residual. The basic steps of the EMD algorithm are:

1) The maximum and minimum values of the original signal $X(t)$. These extreme points are fitted by the curve difference method, and the upper envelope $X_{\max}(t)$ and the lower envelope $X_{\min}(t)$ of the signal are obtained.
2) Average the upper and lower envelope lines by

$$m_1(t) = \frac{X_{\max}(t) + X_{\min}(t)}{2}. \tag{2}$$

3) Subtract the average envelope $m_1(t)$ from the original signal $X(t)$ to obtain the remaining signal $d_1(t)$ as

$$d_1(t) = X(t) - m_1(t). \tag{3}$$

4) Repeat steps 1) to 3) for the remaining signal $d_1(t)$ until it stops when it is less than the screening threshold (SD), So $d_1(t)$ is the first-order modal component $c_1(t)$, i.e. the first IMF. The SD is :

$$SD = \sum_{t=0}^{T} \frac{|d_{k-1}(t) - d_k(t)|^2}{d_{k-1}^2(t)}. \tag{4}$$

5) The first order residual $r_1(t)$ is obtained as the difference of signals $c_1(t)$ and $X(t)$. Replace the original signal $X(t)$ with $r_1(t)$ for steps 1) to 4). After repeating N times, the n-th order modal function $c_n(t)$ and the final standard residual $r_n(t)$ can be obtained. The expression of original signal $X(t)$ after EMD is

$$X(t) = \sum_{1}^{n} c_n(t) + r_n(t). \tag{5}$$

Then, the IMF component is transformed by the Hilbert transformation. In this case, the original signal can be expressed as

$$H(t) = \mathrm{Re}\left(\sum_{q=1}^{Q} a_q(t) \exp\left(i \int w_q(t) dt\right)\right), \tag{6}$$

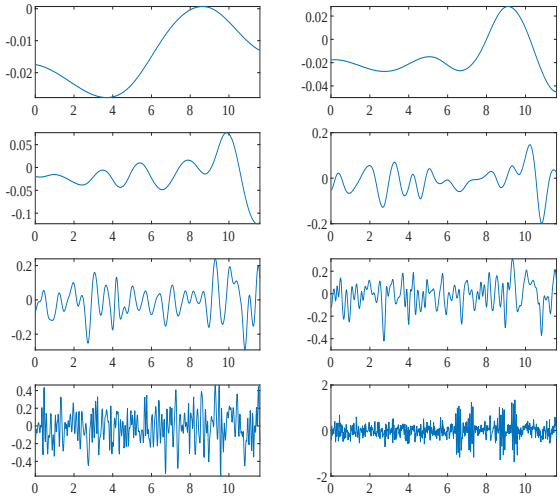

Fig. 3. EMD decomposition of ADS-B signal.

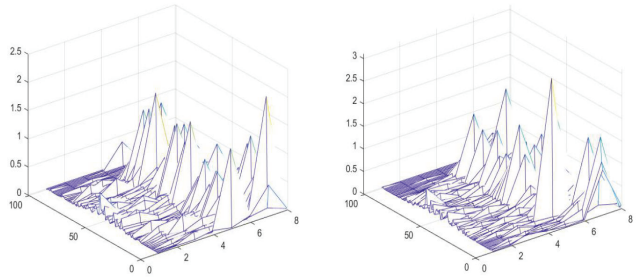

Fig. 4. Hilbert spectrum of different ADS-B signals.

where Re is for the real part, Q stands for the total number of IMF, $a_q(t)$ is instantaneous amplitude and $w_q(t)$ represents the instantaneous frequency of each IMF component. Finally, we calculate the Hilbert spectrum based on all the IMF components.

Fig.3 shows the IMF components of the original signal and Fig.4 shows the Hilbert spectrum of IMF components. From Fig.4, we can also see that the HHT of different ADS-B signals is different.

B. Short-time Fourier Transform

Short-time Fourier transforms (STFT) is related to Fourier transform, which is used to determine the frequency and phase of a sine wave in the local area of a time-varying signal. It is also called the windowed Fourier transform. The time window makes the signal effective only in a certain cell to avoid the deficiency of traditional Fourier transform.

Generally speaking, shorter window can provide better time resolution and longer window can provide better frequency-domain resolution. According to the Heisenberg uncertainty criterion, the time-domain and frequency-domain resolutions of short-time Fourier transform can not reach the optimum, which limits its application. Because the focus of this paper

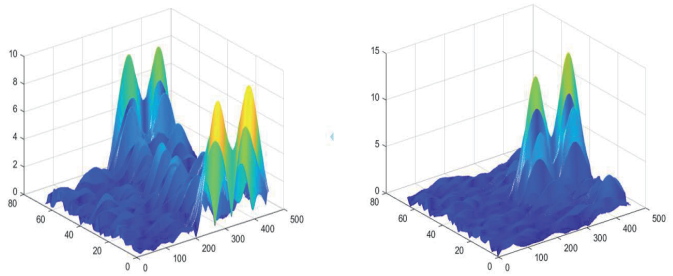

Fig. 5. Short-time Fourier transform of different ADS-B signals

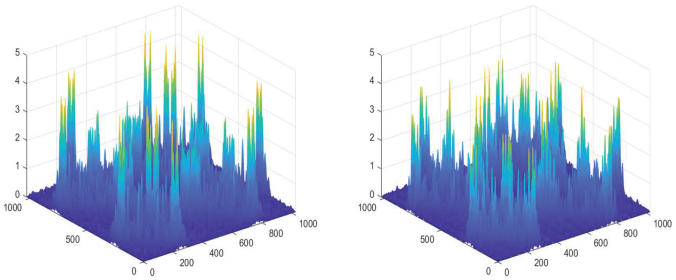

Fig. 6. Bi-spectrum transform of different ADS-B signals.

is the characteristic curve which can reflect the individual difference of fingerprint information, as long as we take a certain time resolution while ensuring enough frequency resolution, we can eliminate the defect of using FFT. STFT can be expressed as

$$X_n(e^{j\varpi}) = \sum_{m=-\infty}^{\infty} X(m)w(n-m)e^{j\varpi m}, \tag{7}$$

where $\omega(n-m)$ is a window function sequence. Different window function sequences will give different Fourier transforms. Short-time Fourier transform has two independent variables n and ω as discrete functions about time n and continuous functions about angular frequency respectively. If we make $\omega = 2\pi nk/N$, then we obtain the discrete short-time Fourier transform, which is sampled in the frequency domain. For continuous functions, STFT can be expressed as

$$STFT\{X(t)\}(\tau, f) = X(\tau, f) = \int_{-\infty}^{\infty} X(t)f(t-\tau)e^{-j2\pi ft}dt. \tag{8}$$

Fig.5 shows the STFT of different ADS-B signals, using the Hamming window with a window length of 64 for the STFT.

C. Bi-spectrum Analysis

The bi-spectrum of the received signal $r(n)$ is estimated by nonparametric method. $r(n)$is divided into Γ segments, each segment contains Δ samples. Then the third-order cyclic cumulant of the received signal $r(n)$ can be written as:

$$\widehat{c}_{3r}(\tau_1, \tau_2) = \frac{1}{\Gamma}\sum_{\gamma=1}^{\Gamma} \chi^{\gamma}(\tau_1, \tau_2), \tag{9}$$

where $\chi^{\gamma}(\tau_1, \tau_2)$ is the third-order cyclic cumulant of each signal. Bi-spectrum estimation of signal $r(n)$ can be expressed as

$$\widehat{B}(w_1, w_2) = \sum_{\tau_1=-\delta}^{\delta}\sum_{\tau_2=-\delta}^{\delta} \hat{c}_{3r}(\tau_1, \tau_2)w(\tau_1, \tau_2)e^{-j(w_1\tau_1+w_2\tau_2)}, \tag{10}$$

where $\delta < \Delta - 1$, and $\omega(\tau_1, \tau_2)$ is the hexagonal window function.The objective function of bi-spectrum dimension reduction can defined as

$$\min_{U_1,\ldots,U_M} \frac{\sum_{i,j}||U_i - U_j||^2 s_{ij}}{\sum_{i,j}||U_i - U_j||^2 d_{ij}}, \tag{11}$$

where $\{\mathrm{U}_1,\ldots,\mathrm{U}_\mathrm{M}\}$ is the compressed bi-spectrum projection space of the original selected bi-spectrum set $\{\mathrm{V}_1,\ldots\mathrm{V}_\mathrm{M}\}$. S and D are the weight matrix of the bi-spectrum, from the same or different emitter. If the emitter are different, $s_{ij} = 0$, otherwise

$$s_{ij} = \frac{1}{\varepsilon}e^{-||V_i - V_j||^2}. \tag{12}$$

If the emitter are the same, $d_{ij} = 0$, otherwise

$$d_{ij} = \frac{1}{\varepsilon}e^{-||V_i - V_j||^2}, \tag{13}$$

where ε is a positive real number.

Then we rewritten (11) as

$$\min_{\omega} \frac{\omega^T V(\varphi_1 \otimes I_n)V^T\omega}{\omega^T V(\varphi_2 \otimes I_n)V^T\omega}, \tag{14}$$

where $\varphi_1 = O - S, \varphi_2 = F - D$, $o_{ii} = \sum_j s_{ij}, f_{ii} = \sum_j d_{ij}$, $\otimes$ is the matrix multiplication. Let $w^T V(\varphi_2 \otimes I_n)V^T w = \delta$, where δ is a nonzero constant, then the Lagrangian function can be defined as

$$\begin{aligned} G(\omega, \lambda) = {} & \omega^T V(\varphi_1 \otimes I_n)V^T\omega \\ & +\lambda(u - \omega^T V(\varphi_2 \otimes I_n)V^T\omega), \end{aligned} \tag{15}$$

where λ is a weight coefficient. To minimize (14), we set $\frac{\partial G(\omega,\lambda)}{\partial w} = 0$, then we can obtain

$$V(\varphi_1 \otimes I_n)V^T w = \lambda V(\varphi_2 \otimes I_n)V^T w. \tag{16}$$

We can calculate the projection matrix $W = [\omega_1, \ldots, w_d]$. The compressed bi-spectrum U_i can be calculated as $U_i = V_i W$.

Fig.6 shows the bi-spectrum transformation diagram of ADS-B signal.

D. Ambiguity Function

The ambiguity function can be used to analyze and design various signals, and can also be used to identify different signal types. Assuming that the delay difference is τ and the frequency shift is μ, the ambiguity function can be defined as

$$A(\tau, \mu) = \int_{-\infty}^{+\infty} X_1(t + \frac{\tau}{2})X_2^*(t - \frac{\tau}{2})e^{j2\pi\mu\tau}dt. \tag{17}$$

Normally, the signal pulse envelope is not rectangular, and the pulse of different radiation source individuals has

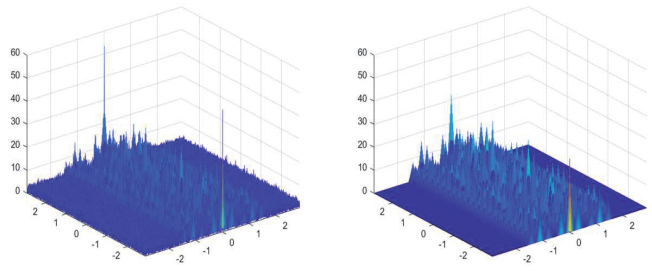

Fig. 7. Ambiguity function of different ADS-B signals.

unique characteristics such as front edge, back edge, and top drop, etc. It is difficult to achieve a good classification effect if the envelope-extracting algorithm is adopted based on the envelope-extracting features alone because the envelope-extracting features are easily affected by noise. Individual characteristics are influenced by many factors. In addition to preserving the envelope characteristics of the signal, the ambiguity function performs very effective center alignment processing on the data, thus eliminating the error caused by pulse envelope misalignment.

The ambiguity functions of different ADS-B signals are shown in Fig.7.

E. Fingerprint Feature Fusion

In the above, the advantages and disadvantages of the above four feature extraction methods has been discussed. we combine these features. Since the RGB image itself is a three-channel image, the combination of feature images will undoubtedly increase the training complexity of the neural network, and the results are not good. In this paper, we make linear combination of the original feature matrices after feature extraction to obtain a new feature map. Since the matrix dimensions obtained by each feature extraction method are different, it is necessary to carry out the pre-processing operation and fix each feature to a size of 200*200. When the feature size is too small, a zero supplementation method is adopted. When the feature size is too large, lower sampling is used for deletion. Finally, each signal is characterized by 4*200*200, and the feature fusion is shown in Fig.8.

IV. Radio Frequency Fingerprint Collaborative Identification Based on Incremental Learning

In machine learning, shallow learning has a high requirement for feature extraction, and feature extraction has great impact on accuracy. Moreover, their recognition accuracy in low signal-to-noise ratio (SNR) is not satisfactory. Deep learning can fully utilize the potential features among the data and has less stringent requirements on feature extraction [37]. Convolutional neural network (CNN) is a deep feed-forward artificial neural network which has been successfully applied to various kinds of data recognition. Alex-Net is used as the recognition architecture in this paper, and the network structure is shown in Fig.9. Alex-Net has 60 million parameters and 65,000 neurons. The first layer is five convolutional layers, and the size and number of convolution kernels in each layer

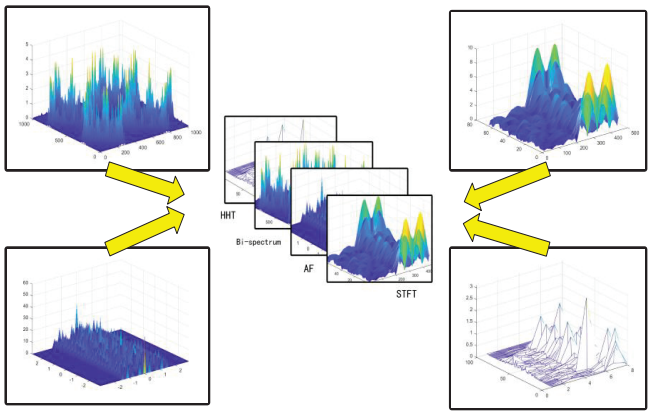

Fig. 8. Fingerprint Feature Fusion.

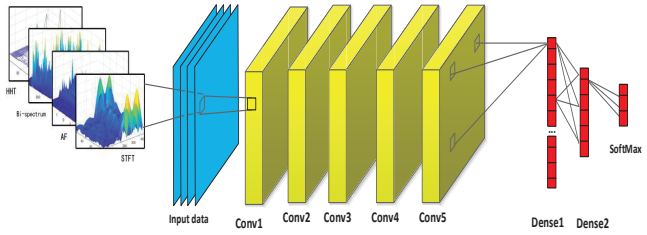

Fig. 9. Network structure for radio frequency fingerprint recognition

are 96×11×11, 256×5×5, 384×3×3, 384×3×3, 256×3×3. A pooling layer is added after each convolutional layer for data compression and computation reduction, both of which are 3*3. The first two layers are 4096 nodes, and the third layer is 1000 nodes. Since the output classification result of this paper is the number of radiation source individuals, the soft-max function with the number of output nodes M (M is the number of radiation source individuals to be identified) is added as the output layer at the end. A drop-out layer is added after each full connection layer to prevent over-fitting. The combined features were input into the network for feature extraction and finally classified by soft-max function. The convolutional neural network architecture for radio frequency fingerprint recognition in this paper is shown in Fig.9.

A. Incremental Classifier and Representation Learning

Traditional training must have all data in advance and train uniformly. But in reality, data is often streamed. Then it is necessary to use the idea of incremental learning. Combine part of the old data and new data into a new data set to train the network. In this way, the training time is greatly reduced and the accuracy is guaranteed. At present, the main challenge of incremental learning is catastrophic forgetting problem (that is, with the increase of training categories, the network fit more to new data, and gradually forget the previous data) [31]. Our method uses some old data to extract features, constructs a unique loss function, and other methods to reduce the catastrophic forgetting problem. The specific steps of this method are as following and the process are shown in Fig.10.

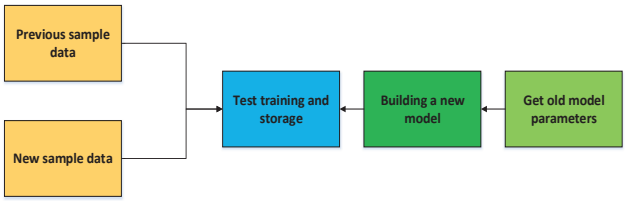

Fig. 10. Incremental classifier and representation learning process

1) First, using the training data of the current category and the data of the old category, train and update the parameter θ(θ is the weight and bias value).

2) Determine the number of data m that can be retained in each category, delete the old category sample set, and retain a small number of samples.

3) A new sample set is constructed by merging the samples of the current category with some of the reserved samples. Record the reconstructed sample set as $P = \{P_1, P_2 ... P_t\}$, then use the feature extractor $\varphi(\cdot)$ to extract the features for each category.

4) Finally, we optimize the classification results according to the following loss function

$$l(\Theta) = -\sum_{(x_i,y_i)\in D}\{\sum_{y=s}^{t}\delta_{y=y_i}\log\, g_y(x_i)+\delta_{y\neq y_i}\log(1-g_y(x_i)) + \sum_{y=1}^{s-1} q_i^y \log\, g_y(x_i) + (1-q_i^y)\log(1-g_y(x_i))\} \tag{18}$$

where

$$g_y(x_i) = \frac{1}{1+e^{-w_y^T\phi(z)}}, \tag{19}$$

and

$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}. \tag{20}$$

By setting a T value greater than 1, it will aggravate the consequences of training errors, which is equivalent to weight training, so that the training accuracy is higher. The new loss function is much better than the cross-entropy function. Algorithm 1 summarizes the training process of incremental classifier and representation learning.

B. Learning Without Forgotten

Considering the method mentioned in [35] for incremental training of data. Fig.11 shows the training process of learning without forgotten.

Firstly, the network is trained based on the original data, and obtain the parameters of the original task, among which the shared parameter θ_s(feature extraction layer). The old task parameter θ_0 is some parameter information of the old network (such as weight, learning rate, bias, etc). Then we input the data of the new category into the original network, record the response record y_o of the original network (for image classification, the response is naturally the set of label probabilities of each training image), then add the nodes of each new category to the output layer, fully connect to the following layer, and randomly initialize the weight as θ_n. Then the network is trained to minimize the loss of all tasks using the following loss functions,

$$\arg\min_{\hat{\theta}_s,\hat{\theta}_0,\hat{\theta}_n}(\lambda_0 L_{old}(Y_0,\hat{Y}_0) + L_{new}(Y_n,\hat{Y}_n)), \tag{21}$$

where

$$L_{new}(y_n,\hat{y}_n) = -y_n * \log\hat{y}_n, \tag{22}$$

and $\hat{y}_n$ is the output value using softmax function, y_n is the output tag value.

$$\begin{aligned} L_{old}(y_0,\hat{y}_0) &= -H(y_0^{'},\hat{y}_0^{'}) \\ &= -\sum_{i=1}^{l} y_0^{'}\log\hat{y}_0^{'(i)}, \end{aligned} \tag{23}$$

where l is the total number of labels, and

$$y_0^{'(i)} = \frac{(y_0^{(i)})^{\frac{1}{T}}}{\sum_j (y_0^{(j)})^{\frac{1}{T}}}, \hat{y}_0^{'(i)} = \frac{(\hat{y}_0^{(i)})^{\frac{1}{T}}}{\sum_j (\hat{y}_0^{(j)})^{\frac{1}{T}}}. \tag{24}$$

Compared with the use of old data, it can achieve faster training speed without using the old data at all, and it does not need the old data, which also saves the storage space. But the problem is recognition accuracy of the old category is not good, and if the new category is smaller than the old category, the recognition error will be significantly increased. In this method, if we use some old data, the recognition accuracy will

Algorithm 1 Training process of incremental classifier and representation learning

1: Train a neural network in advance to save model parameters and other information;
2: Select the old samples randomly, take out some samples and new types of data to form new training data;
3: Using feature extractor $\varphi(\cdot)$ to extract features from old and new data (only a part of the old data) and obtain the predicted value;
4: Put the predicted value into $l(\Theta) = -\sum_{(x_i,y_i)\in D}\{\sum_{y=s}^{t}\delta_{y=y_i}\log\, g_y(x_i)+\delta_{y\neq y_i}\log(1-g_y(x_i)) + \sum_{y=1}^{s-1} q_i^y \log\, g_y(x_i) + (1-q_i^y)\log(1-g_y(x_i))\}$ for optimization, and finally obtain the model.

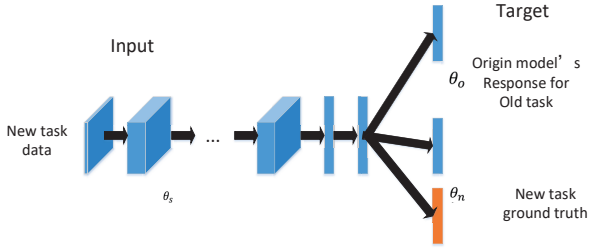

Fig. 11. Training process of learning without forgotten

be improved. Algorithm 2 summarizes the training process of learning without forgotten.

Algorithm 2 Training process of learning without forgotten

1: obtain the training results of the old categories through training, and obtain θ_{s} and θ_0;
2: Add new classification node to output layer, initialize parameter θ_{n} randomly, freeze $\theta_{\mathrm{s}}$and θ_0 to train to θ_{n} convergence;
3: $\theta_{\mathrm{n}}, \theta_{\mathrm{s}}$ and θ_0 are trained together to convergence, and the new task model is obtained and $\underset{\hat{\theta}_s,\hat{\theta}_0,\hat{\theta}_n}{\arg\min}(\lambda_0 L_{old}(Y_0,\hat{Y}_0)+ L_{new}(Y_n,\hat{Y}_n))$ is used to optimize its convergence.

C. Collaborative Radio Frequency Fingerprint Identification

In the actual electronic countermeasure environment, the system is usually designed as a distributed sensor system in order to obtain all dimensions of electromagnetic information. For the distributed sensor system composed of neural network, each sub network is faced with different electromagnetic environment, and the recognition effect is also different. The decision made for different data is likely to be contrary to the fact, which will lead to the decline of the accuracy of individual emitter recognition. In view of this, an individual emitter recognition method based on decision fusion is proposed [36].

In the distributed emitter identification system, the decision result set of each sub recognition system is $\{\sigma_0, \sigma_1\}$, where σ_1 represents the untrained radiation source and σ_0 represents the known radiation source. And the basic probability assignment $M_i : \{m_i(\sigma_0), m_i(\sigma_1)\}$ of each subsystem to the emitter individual is obtained. The Jousselme distance of the evidence obtained by each sensor node is calculated by

$$dis_j = \sqrt{\frac{1}{2}(M_i - M_j)^T D\,(M_i - M_j)}. \tag{25}$$

Then the evidence distance matrix of each subsystem can be obtained by

$$D^{\Omega} = \begin{bmatrix} 0 & dis_{j12} & \cdots & dis_{j1n} \\ dis_{j21} & 0 & \cdots & dis_{j2n} \\ \vdots & \vdots & \ddots & \vdots \\ dis_{jn1} & dis_{jn2} & \cdots & dis_{jnn}, \end{bmatrix} \tag{26}$$

where dis_{jmn} represents the Jousselme distance between the evidence obtained by the n-sub-recognition system and the m-sub-recognition system, which represents the conflict degree between the evidences obtained. According to the conflict degree between the evidences obtained by each subsystem, the trust degree of the recognition results obtained by each subsystem can be deduced as

$$\alpha_{ij} = 1 - dis_{jij}. \tag{27}$$

It can be concluded that the reliability of the recognition result of the i-th subsystem to the recognition result of the j-th subsystem is

$$Sup(M_i) = \sum_{j=1, j\neq i}^{n} \alpha_{ij}. \tag{28}$$

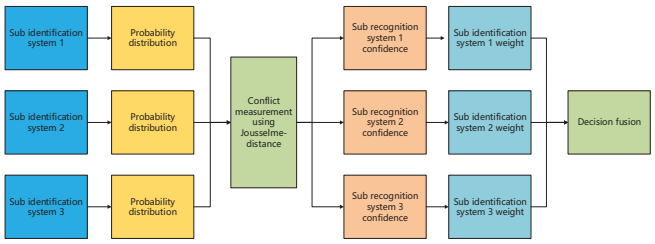

Fig. 12. Collaborative identification based on Jousselme evidence distance

After normalizing the credibility, the weight of recognition results of each subsystem is

$$w_i = \frac{Sup(M_i)}{\sum_{i=1}^{n} Sup(M_i)}, \tag{29}$$

then the recognition result of the$i$-th subsystem is modified to

$$m(A) = \sum_{i=1}^{n} w_i \times m_i(A). \tag{30}$$

The above collaborative identification method based on Jousselme evidence distance is shown in Fig.12.

V. Numerical Results and Discussion

The simulation experiments was performed on an Intel Core i9-9920x desktop computer with a 3.50ghz CPU and 96GB of RAM, using two RTX2080Ti graphics cards. MATLAB version is R2018b, TensorFlow version is GPU-1.14. The simulation signal contains fingerprint information such as frequency offset, phase distortion, and harmonic distortion. The simulation signal uses an S-mode transponder to expand the message (1090es, The ADS-B signal of 1090 MHz mode s extended squitter) has a working frequency of 1090 MHz, a data rate of 1 Mbps, and a modulation mode of PPM and 2ASK. The period of the signal is 120 μs, the leading pulse lasts 8 μs, and the duration is 0.5 μs. The starting time of the pulses is at 0.1 μs, 3.5 μs, and 4.5 μs respectively. The information pulse occupies 112 μs and transmits 112 bits of data. One bit of data represents a message, including the position, altitude, speed, heading, identification number, and other information of the aircraft. The 01 and 10 binary data used after PPM modulation represents each message. We intercept the data of 5 μs length of ADS-B signal's leading pulse, use the frequency of 600MHz to sample, set the signal-to-noise ratio range between -5 dB and 5 dB, and set three different individuals in total, each of which generates five different signals. Table I shows the target parameter settings.

Fig.13 shows the recognition accuracy of the four feature extraction methods in the Gaussian channel. From Fig.13, it can be seen that feature extraction has great impact on the recognition rate. The recognition rate of bi-spectrum transform and ambiguity function can reach 94%, while the recognition rate of Hilbert-Huang transforms and short-time Fourier transform is worse. Fig.14 shows the recognition accuracy of four feature extraction methods in Rayleigh fading channel. From

TABLE I
TARGET PARAMETER SETTINGS.

Individual	Parameter	Parameter Settings				
Target 1	Modulation frequency of the phase noise(MHz)	4	6	7	10	15
	Phase modulation coefficient	0.16	0.27	0.32	0.15	0.25
	Harmonic component	1	0.5	0.3	0.2	0.1
Target 2	Modulation frequency of the phase noise(MHz)	2	5	9	11	13
	Phase modulation coefficient	0.21	0.32	0.15	0.24	0.28
	Harmonic component	1	0.8	0.6	0.4	0.2
Target 3	Modulation frequency of the phase noise(MHz)	3	5	6	8	12
	Phase modulation coefficient	0.34	0.3	0.23	0.21	0.26
	Harmonic component	1	0.1	0.08	0.05	0.03

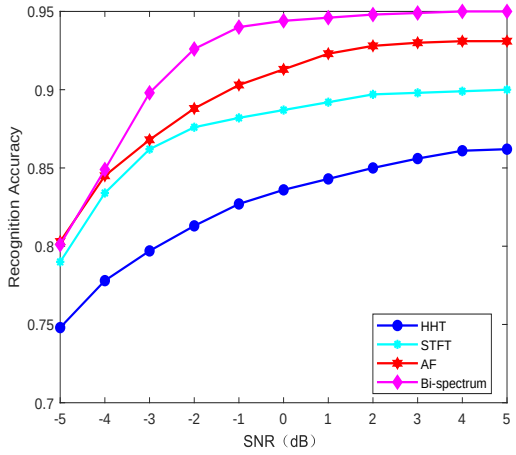

Fig. 13. Recognition performance of four methods with different SNRs over Gaussian channel.

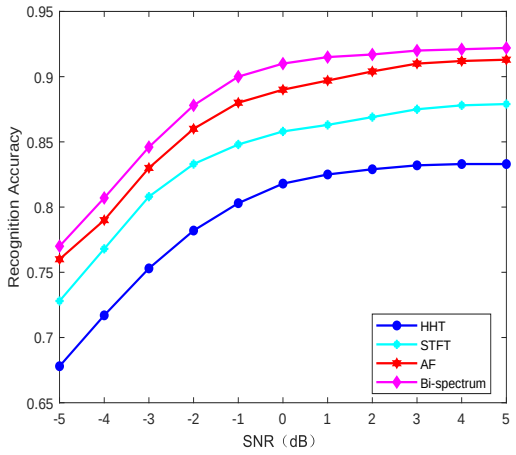

Fig. 14. Recognition performance of four methods with different SNRs over Rayleigh channel.

Fig.14, it can be seen that the recognition performance of the four methods have decreased over the Rayleigh fading channel.

Fig.15 shows the recognition accuracy for different SNRs combining different feature extraction. As can be seen, after the feature combination, the information among the features is complementary to each other, which makes the recognition accuracy bigger than that of a single feature as input, and also improves the recognition accuracy in the case of low SNR. Fig.16 shows the recognition accuracy in Rayleigh fading channel. The recognition accuracy of the Rayleigh fading channel is smaller than the Gaussian channel.

Fig.17 shows the RF-FI identification accuracy curve of the radiation source proposed in this paper under different convolutional neural network architectures and Table II shows the training time of different Networks. As can be seen, the VGG-16 network has the highest recognition accuracy, but it is not an ideal training method due to its deep network structure, long training time, and its loss function is not easy to converge under the deep network. The recognition accuracy of LeNet-5 and ordinary convolutional neural networks are significantly lower than that of Alex-Net. The Alex-Net network has good recognition accuracy, and its accuracy in the validation set is up to 95%. The existing identification methods mostly use the deep neural networks, such as the Dense-net, which reaches 90 layers, and knows that the time and storage space spent training these networks is quite amazing. Based on the Alex-net network, there are only five layer of convolution and some full connectivity, and the complexity of the model is greatly reduced. In this paper, the computational complexity is as follows: the complexity of feature extraction is $O(M \log M)$, where M is the data length. The complexity of the network is $O\left(\sum_{l=1}^{N} S_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l\right)$, where l represents the number of convolution layers, S is the length of the convolution kernel's output feature map, K is the length of each convolution kernel, and C is the number of channels per convolution kernel and N represents the depth of the network.

Fig.18 shows the recognition accuracy under different training strategies. The traditional strategy, where all the data are put together for training, works the best, and incremental learning leads to loss of accuracy. However, as shown in Table III, due to the joint training of traditional training requirements and all data, the storage space needs to be improved, and the training time greatly exceeds the incremental learning method, and the incremental learning greatly reduces the training time and storage space needs on the premise of sacrificing a certain accuracy. The recognition rate is better than 90% when the SNR is greater than 0 dB. The figure shows that the method is effective and practical.

Fig. 19 shows shows the change of recognition accura-

TABLE II
TRAINING TIME OF THE DIFFERENT NETWORKS.

Network	VGG-16	AlexNet	Ordinary CNN	LeNet-5
Training Time (s)	8536	3558	3072	2598

TABLE III
TRAINING TIME OF THE DIFFERENT TRAINING METHODS.

Training methods	Traditional method	method 1 with 1/4 old data	method 1 with 1/10 old data	method 2
Training Time (s)	9206	4383	3558	3290

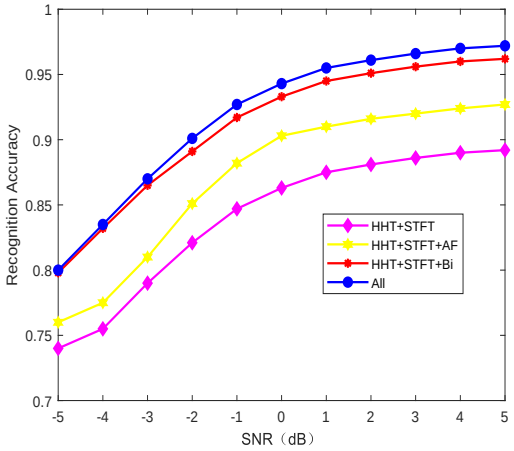

Fig. 15. Recognition performance of different feature fusion methods with different SNRs over Gaussian channel

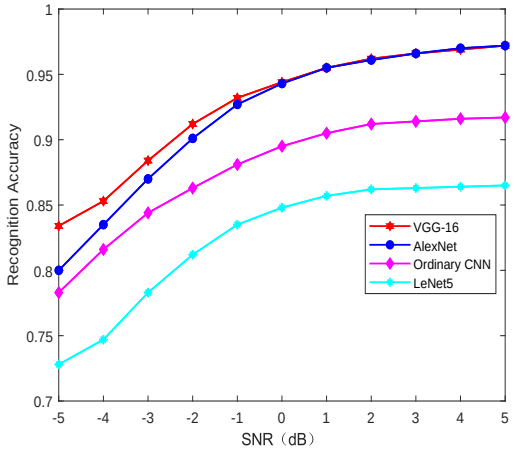

Fig. 17. Recognition performance with different network architectures

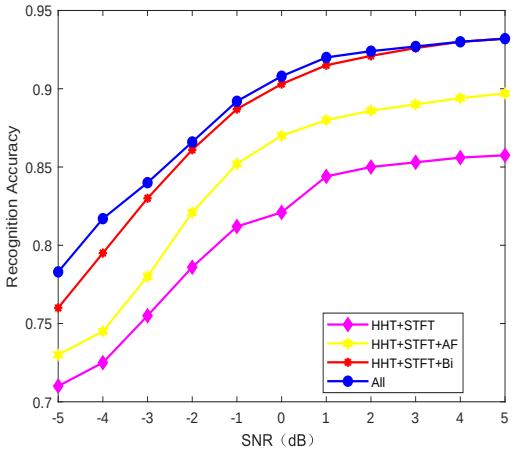

Fig. 16. Recognition performance of different feature fusion methods with different SNRs over Rayleigh channel

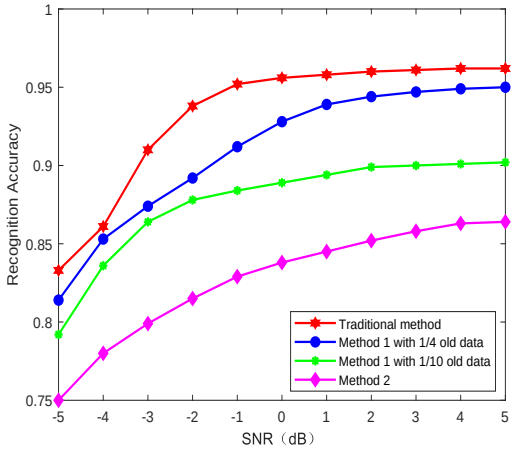

Fig. 18. Recognition performance by different training methods

cy with SNR under different Doppler frequency shifts and channel delays. From Fig. 19, we can see that the proposed method is robust to Doppler frequency shifts and channel delays when SNR over 0dB. Fig. 20 shows the recognition performance with the number of individual targets under the different SNRs and channels. From Fig. 20, we can see that recognition performance decreases slightly as the number of individual targets increases.

VI. CONCLUSION

In this paper, we proposed a RF-FI framework based on incremental learning to solve the problem of blind signals individual identification of the distributed system. Firstly, Hilbert Huang transforms, short-time Fourier transform, bi-spectrum transform, ambiguity function methods are used to extract fingerprint information, fully reflecting the characteristics of the signal. Then, these fingerprint information are fused and input into the convolution neural network for intelligent processing, thereby realizing RF-FI identification of the radiation source and decision fusion method is applied to solve the problem

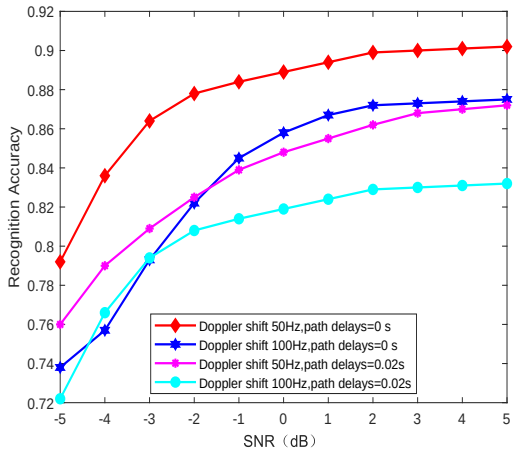

Fig. 19. Recognition performance with different Doppler frequency shift and channel delay.

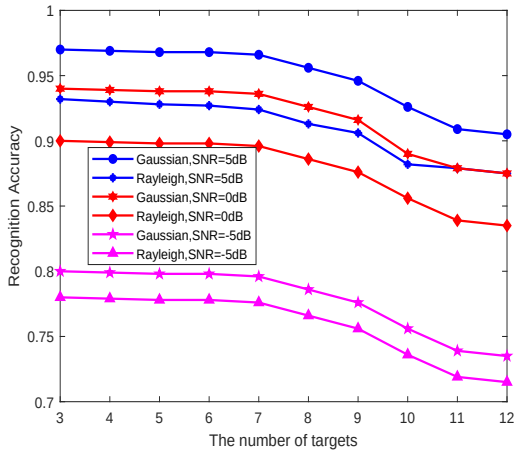

Fig. 20. Recognition performance with the number of individual targets under the different SNRs and channels.

of inconsistent recognition accuracy in distributed system. When newly arrived untrained data arrives, we use incremental learning to train these data which reduces the training time and storage space. Finally, a lot of simulation studies are carried out to verify the effectiveness of the proposed method. Simulation results show that the framework and method can shorten the training time, reduce the space needed for data storage, and obtain good recognition accuracy under low SNR.

REFERENCES

[1] G. Gok, Y. K. Alp and O. Arikan, "A New Method for Specific Emitter Identification With Results on Real Radar Measurements," *IEEE Transactions on Information Forensics and Security,* vol. 15, pp. 3335-3346, 2020.

[2] J. Aliasgari, M. Forouzandeh and N. Karmakar, "Chipless RFID Readers for Frequency-Coded Tags: Time-Domain or Frequency-Domain," *IEEE Journal of Radio Frequency Identification,* vol. 4, no. 2, pp. 146-158, June 2020.

[3] K. Shah and Z. Narmavala, "A Survey on Green Internet of Things," *2018 Fourteenth International Conference on Information Processing (ICINPRO), Bangalore,* India, 2018, pp. 1-4.

[4] L. Roselli et al., "Review of the present technologies concurrently contributing to the implementation of the Internet of Things (IoT) paradigm: RFID, Green Electronics, WPT and Energy Harvesting," *2015 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet),* San Diego, CA, 2015, pp. 1-3, doi: 10.1109.

[5] M. Zhu, X. Zhang, Y. Qi and H. Ji, "Compressed Sensing Mask Feature in Time-Frequency Domain for Civil Flight Radar Emitter Recognition," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Calgary, AB, 2018, pp. 2146-2150.

[6] K. Li, "Radar Emitter Identification Based on Improved Convolutional Neural Network," *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China,* 2019, pp. 118-121.

[7] H. Yang et al., "Intelligent Reflecting Surface Assisted Anti-Jamming Communications: A Fast Reinforcement Learning Approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1963-1974, March 2021.

[8] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao and Q. Wu, "Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Secure Wireless Communications," *IEEE Transactions on Wireless Communications,* vol. 20, no. 1, pp. 375-388, Jan. 2021.

[9] S. Fan, H. Zhang, Y. Zeng and W. Cai, "Hybrid Blockchain-Based Resource Trading System for Federated Learning in Edge Computing," *IEEE Internet of Things Journal,* vol. 8, no. 4, pp. 2252-2264, 15 Feb.15, 2021.

[10] W. Y. B. Lim et al., "Hierarchical Incentive Mechanism Design for Federated Machine Learning in Mobile Networks," *IEEE Internet of Things Journal,* vol. 7, no. 10, pp. 9575-9588, Oct. 2020.

[11] S. Guo, R. E. White and M. Low, "A comparison study of radar emitter identification based on signal transients," *2018 IEEE Radar Conference (RadarConf18), Oklahoma City, OK,* 2018, pp. 0286-0291,

[12] Chunyun Song, Jianmin Xu and Yi Zhan, "A method for specific emitter identification based on empirical mode decomposition," *2010 IEEE International Conference on Wireless Communications, Networking and Information Security, Beijing,* 2010, pp. 54-57.

[13] Y. Liu, H. An and S. Bian, "Hilbert-Huang Transform and the Application," 2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS), Dalian, China, 2020, pp. 534-539.

[14] X.D. Zhang, Y. Shi, and Z. Bao, "A new feature vector using selected bispectra for signal classification with application in radar target recognition,"*IEEE Trans. Signal Process.,* vol. 49, no. 9, pp. 1875-1885, Sep. 2001.

[15] R. Cao and J. Cao, "Radar Emitter Identification with Bispectrum based LBP and Extreme Learning Machine," *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), Shanghai, China,* 2018, pp. 1-5.

[16] N. Kang, M. He, J. Han and B. Wang, "Radar emitter fingerprint recognition based on bispectrum and SURF feature," *2016 CIE International Conference on Radar (RADAR), Guangzhou,* 2016, pp. 1-5.

[17] Tao-wei Chen, Wei-dong Jin and Jie Li, "Feature extraction using surrounding-line integral bispectrum for radar emitter signal," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong,* 2008, pp. 294-298.

[18] Y. Li, W. Peng, X. Zhu, G. Wei, H. Yu and G. Qian, "Based on short-time fourier transform impulse frequency response analysis in the application of the transformer winding deformation," *2016 IEEE International Conference on Mechatronics and Automation, Harbin,* 2016, pp. 371-375.

[19] G. Lopez-Risueo, J. Grajal, and A. Sanz-Osorio, "Digital channelized receiver based on time-frequency analysis for signal interception,"*IEEE Trans. Aerosp. Electron. Syst.,* vol. 41, no.3, pp.879-898, Jul. 2005.

[20] K. Merchant, S. Revay, G. Stantchev and B. Nousain, "Deep Learning for RF Device Fingerprinting in Cognitive Communication Networks," *IEEE Journal of Selected Topics in Signal Processing,* vol. 12, no. 1, pp. 160-167, Feb. 2018.

[21] L. Peng, A. Hu, J. Zhang, Y. Jiang, J. Yu and Y. Yan, "Design of a Hybrid RF Fingerprint Extraction and Device Classification Scheme," *IEEE Internet of Things Journal,* vol. 6, no. 1, pp. 349-360, Feb. 2019.

[22] L. Peng, J. Zhang, M. Liu and A. Hu, "Deep Learning Based RF Fingerprint Identification Using Differential Constellation Trace Figure," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 1, pp. 1091-1095, Jan. 2020.

[23] F. Arlery, R. Kassab, U. Tan and F. Lehmann, "Efficient optimization of the ambiguity functions of multi-static radar waveforms," *2016 17th International Radar Symposium (IRS), Krakow,* 2016, pp. 1-6.

[24] A. Aubry, A. Bazzoni, V. Carotenuto, A. De Maio and P. Failla, "Cumulants-based Radar Specific Emitter Identification," *2011 IEEE International Workshop on Information Forensics and Security, Iguacu Falls,* 2011, pp. 1-6.

[25] M. Liu, G. Liao, N. Zhao, H. Song and F. Gong. “Data-driven deep learning for dignal classification in industrial cognitive radio networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3412-3421, May 2021.

[26] M. Liu, K. Yang, N. Zhao, Y. Chen, H. Song, F. Gong, “Intelligent signal classification in industrial distributed wireless sensor networks-based IIoT," *IEEE Transactions on Industrial Informatics*, 17(7): 4946 - 4956, July 2021.

[27] S. Hou, X. Pan, C. C. Loy, Z. Wang and D. Lin, "Learning a Unified Classifier Incrementally via Rebalancing," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA,* 2019, pp. 831-839.

[28] K. Nguyen, C. Fookes and S. Sridharan, "Improving deep convolutional neural networks with unsupervised feature learning," *2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC,* 2015, pp. 2270-2274.

[29] D. Hu, F. Nie and X. Li, "Deep Multimodal Clustering for Unsupervised Audiovisual Learning," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA,* 2019, pp. 9240-9249.

[30] R. de Lemos and M. Grze, "Self-Adaptive Artificial Intelligence," *2019 IEEE/ACM 14th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), Montreal, QC, Canada,* 2019, pp. 155-156.

[31] S. Rebuffi, A. Kolesnikov, G. Sperl and C. H. Lampert, "iCaRL: Incremental Classifier and Representation Learning," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI,* 2017, pp. 5533-5542.

[32] S. Hou, X. Pan, C. C. Loy, Z. Wang and D. Lin, "Learning a Unified Classifier Incrementally via Rebalancing," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA,* 2019, pp. 831-839.

[33] F. Cermelli, M. Mancini, S. Rota Bulo, E. Ricci and B. Caputo, "Modeling the Background for Incremental Learning in Semantic Segmentation," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA,* 2020, pp. 9230-9239.

[34] S. Dang, Z. Cao, Z. Cui, Y. Pi and N. Liu, "Open Set Incremental Learning for Automatic Target Recognition," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 57, no. 7, pp. 4445-4456, July 2019.

[35] Z. Li and D. Hoiem, "Learning without Forgetting," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 12, pp. 2935-2947, 1 Dec. 2018.

[36] L. Sun, Y. Zhang, Z. Fu, G. Zheng, Z. He and J. Pu, "An Approach to Multi-Sensor Decision Fusion Based on the Improved Jousselme Evidence Distance," *2018 International Conference on Control, Automation and Information Sciences (ICCAIS),* 2018, pp. 189-193.

[37] Q. Lao, X. Jiang, M. Havaei and Y. Bengio, "A Two-Stream Continual Learning System With Variational Domain-Agnostic Feature Replay,"*IEEE Transactions on Neural Networks and Learning Systems*, 2021, doi: 10.1109/TNNLS.2021.3057453.